PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07K 14/705, 16/28, C12N 15/12 | A1

(11) International Publication Number: WO 96/30406

(43) International Publication Date: 3 October 1996 (03.10.96)

(21) International Application Number: PCT/US95/04079

(22) International Filing Date: 30 March 1995 (30.03.95)

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850-3338 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LI, Yi [CN/US]; 16125 Howard Landing Drive, Gaithersburg, MD 20878 (US). CAO, Liang [CN/GB]; 18B Suncrest Tower, Monmouth Terrace, Hong Kong (HK). NI, Jian [CN/US]; Apartment #204, 305 West Side Drive, Gaithersburg, MD 20878 (US). GENTZ, Reiner [DE/US]; 13404 Fairland Park Drive, Silver Spring, MD 20904 (US). BULT, Carol, J. [US/US]; 334 Park Hall South, Laurel, MD 20824 (US). SUTTON, Granger, G., III [US/US]; 6409 Snowman Court, Columbia, MD 21045 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US).

(74) Agents: OLSTEIN, Elliot, M.; Carella, Byrne, Bain, Gilfillan, Cecchi, Stewart & Olstein, 6 Becker Farm Road, Roseland, NJ 07068 (US) et al.

(81) Designated States: AU, CA, CN, JP, KR, MX, NZ, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: HUMAN G-PROTEIN COUPLED RECEPTORS

(57) Abstract

Human G-protein coupled receptor polypeptides and DNA (RNA) encoding such polypeptides and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing such polypeptides for identifying antagonists and agonists to such polypeptides and methods of using the agonists and antagonists therapeutically to treat conditions related to the underexpression and overexpression of the G-protein coupled receptor polypeptides, respectively. Also disclosed are diagnostic methods for detecting a mutation in the G-protein coupled receptor nucleic acid sequences and an altered level of the soluble form of the receptors.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

HUMAN G-PROTEIN COUPLED RECEPTORS

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides. More particularly, the polypeptides of the present invention are human 7-transmembrane receptors. The transmembrane receptors are G-protein coupled receptors sometimes hereinafter referred to individually as GPR1, GPR2, GPR3 and GPR4. The invention also relates to inhibiting the action of such polypeptides.

It is well established that many medically significant biological processes are mediated by proteins participating in signal transduction pathways that involve G-proteins and/or second messengers, e.g., cAMP (Lefkowitz, Nature, 351:353-354 (1991)). Herein these proteins are referred to as proteins participating in pathways with G-proteins or PPG proteins. Some examples of these proteins include the GPC receptors, such as those for adrenergic agents and dopamine (Kobilka, B.K., et al., PNAS, 84:46-50 (1987); Kobilka, B.K., et al., Science, 238:650-656 (1987); Bunzow, J.R., et al., Nature, 336:783-787 (1988)), G-proteins themselves, effector proteins, e.g., phospholipase C, adenyl cyclase, and phosphodiesterase, and actuator proteins, e.g., protein

kinase A and protein kinase C (Simon, M.I., et al., Science, 252:802-8 (1991)).

For example, in one form of signal transduction, the effect of hormone binding is activation of an enzyme, adenylate cyclase, inside the cell. Enzyme activation by hormones is dependent on the presence of the nucleotide GTP, and GTP also influences hormone binding. A G-protein connects the hormone receptors to adenylate cyclase. G-protein was shown to exchange GTP for bound GDP when activated by hormone receptors. The GTP-carrying form then binds to an activated adenylate cyclase. Hydrolysis of GTP to GDP, catalyzed by the G-protein itself, returns the G-protein to its basal, inactive form. Thus, the G-protein serves a dual role, as an intermediate that relays the signal from receptor to effector, and as a clock that controls the duration of the signal.

The membrane protein gene superfamily of G-protein coupled receptors has been characterized as having seven putative transmembrane domains. The domains are believed to represent transmembrane α-helices connected by extracellular or cytoplasmic loops. G-protein coupled receptors include a wide range of biologically active receptors, such as hormone, viral, growth factor and neuroreceptors.

G-protein coupled receptors have been characterized as including these seven conserved hydrophobic stretches of about 20 to 30 amino acids, connecting at least eight divergent hydrophilic loops. The G-protein family of coupled receptors includes dopamine receptors which bind to neuroleptic drugs used for treating psychotic and neurological disorders. Other examples of members of this family include calcitonin, adrenergic, endothelin, cAMP, adenosine, muscarinic, acetylcholine, serotonin, histamine, thrombin, kinin, follicle stimulating hormone, opsins and rhodopsins, odorant, cytomegalovirus receptors, etc.

Most G-protein coupled receptors have singl conserved cysteine residues in each of the first two extracellular loops which form disulfide bonds that are believed t stabilize functional protein structure. The 7 transmembrane regions are designated as TM1, TM2, TM3, TM4, TM5, TM6, and TM7. TM3 is also implicated in signal transduction.

Phosphorylation and lipidation (palmitylation or farnesylation) of cysteine residues can influence signal transduction of some G-protein coupled receptors. Most G-protein coupled receptors contain potential phosphorylation sites within the third cytoplasmic loop and/or the carboxy terminus. For several G-protein coupled receptors, such as the β-adrenoreceptor, phosphorylation by protein kinase A and/or specific receptor kinases mediates receptor desensitization.

The ligand binding sites of G-protein coupled receptors are believed to comprise a hydrophilic socket formed by several G-protein coupled receptors transmembrane domains, which socket is surrounded by hydrophobic residues of the G-protein coupled receptors. The hydrophilic side of each G-protein coupled receptor transmembrane helix is postulated to face inward and form the polar ligand binding site. TM3 has been implicated in several G-protein coupled receptors as having a ligand binding site, such as including the TM3 aspartate residue. Additionally, TM5 serines, a TM6 asparagine and TM6 or TM7 phenylalanines or tyrosines are also implicated in ligand binding.

G-protein coupled receptors can be intracellularly coupled by heterotrimeric G-proteins to various intracellular enzymes, ion channels and transporters (see, Johnson *et al.*, Endoc., Rev., 10:317-331 (1989)). Different G-protein α-subunits preferentially stimulate particular effectors to modulate various biological functions in a cell. Phosphorylation of cytoplasmic residues of G-protein coupled receptors have been identified as an important mechanism for

the regulation of G-protein coupling of some G-protein coupled rec ptors.

G-protein coupled receptors are found in numerous sites within a mammalian host, for example, dopamine is a critical neurotransmitter in the central nervous system and is a G-protein coupled receptor ligand.

In accordance with one aspect of the present invention, there are provided novel polypeptides which have been putatively identified as G-protein coupled receptors and biologically active and diagnostically or therapeutically useful fragments and derivatives thereof. The polypeptides of the present invention are of human origin.

In accordance with another aspect of the present invention, there are provided isolated nucleic acid molecules encoding human G-protein coupled receptors, including mRNAs, DNAs, cDNAs, genomic DNA as well as antisense analogs thereof and biologically active and diagnostically or therapeutically useful fragments thereof.

In accordance with a further aspect of the present invention, there is provided a process for producing such polypeptides by recombinant techniques which comprises culturing recombinant prokaryotic and/or eukaryotic host cells, containing a human G-protein coupled receptor nucleic acid sequence, under conditions promoting expression of said protein and subsequent recovery of said protein.

In accordance with yet a further aspect of the present invention, there are provided antibodies against such polypeptides.

In accordance with another embodiment, there is provided a process for using the receptors to screen for receptor antagonists and/or agonists and/or receptor ligands.

In accordance with still another embodiment of the present invention there is provided a process of using such agonists to stimulate the G-protein coupled receptors for the

treatment of conditions related to the under-expression of the G-protein coupled receptors.

In accordance with another aspect of the present invention there is provided a process of using such antagonists for inhibiting the action of the G-protein coupled receptors for treating conditions associated with over-expression of the G-protein coupled receptors.

In accordance with yet another aspect of the present invention there is provided non-naturally occurring synthetic, isolated and/or recombinant G-protein coupled receptor polypeptides which are fragments, consensus fragments and/or sequences having conservative amino acid substitutions, of at least one transmembrane domain of the G-protein coupled receptor, such that G-protein coupled receptor polypeptides of the present invention may bind G-protein coupled receptor ligands, or which may also modulate, quantitatively or qualitatively, G-protein coupled receptor ligand binding.

In accordance with still another aspect of the present invention there are provided synthetic or recombinant G-protein coupled receptor polypeptides, conservative substitution and derivatives thereof, antibodies, anti-idiotype antibodies, compositions and methods that can be useful as potential modulators of G-protein coupled receptor function, by binding to ligands or modulating ligand binding, due to their expected biological properties, which may be used in diagnostic, therapeutic and/or research applications.

It is still another object of the present invention to provide synthetic, isolated or recombinant polypeptides which are designed to inhibit or mimic various G-protein coupled receptors or fragments thereof, as receptor types and subtypes.

In accordance with yet a further aspect of the present invention, there is also provided diagnostic probes comprising nucleic acid molecules of sufficient length to

specifically hybridize to the G-protein coupled receptor nucleic acid s quences.

In accordance with yet another object of the present invention, there is provided a diagnostic assay for detecting a disease or susceptibility to a disease related to a mutation in a G-protein coupled receptor nucleic acid sequence.

These and other aspects of the present invention should be apparent to those skilled in the art from the teachings herein.

The following drawings are illustrative of embodiments of the invention and are not meant to limit the scope of the invention as encompassed by the claims.

Figures 1-4 show the cDNA sequences and the corresponding deduced amino acid sequences of the four G-protein coupled receptors of the present invention, respectively. The standard one-letter abbreviation for amino acids are used. Sequencing was performed using a 373 Automated DNA sequencer (Applied Biosystems, Inc.). Sequencing accuracy is predicted to be greater than 97% accurate.

Figure 5 is an illustration of the amino acid homology between GPR1 (top line) and odorant receptor-like protein (bottom line).

Figure 6 illustrates the amino acid homology between GPR2 (top line) and the human Endothelial Differentiation Gene-1 (EDG-1) (bottom line).

Figure 7 illustrates the amino acid homology between GPR3 (top line) and a human G-protein coupled receptor open reading frame (ORF) (bottom line).

Figure 8 illustrates the amino acid homology between GPR4 and the chick orphan G-protein coupled receptor (bottom line).

In accordance with an aspect of the present invention, there are provided isolated nucleic acids (polynucleotides)

which encode for the mature polypeptides having the deduced amino acid sequences of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or for the mature polypeptides encoded by th cDNAs of the clones deposited as ATCC Deposit No. 75981 (GPR1), 75983 (GPR2), 75976 (GPR3), 75979 (GPR4) on December 16, 1994.

A polynucleotide encoding the GPR1 polypeptide of the present invention may be isolated from the human breast. The polynucleotide encoding GPR1 was discovered in a cDNA library derived from human eight-week-old embryo. It is structurally related to the G protein-coupled receptor family. It contains an open reading frame encoding a protein of 296 amino acid residues. The protein exhibits the highest degree of homology to an odorant receptor-like protein with 66 % identity and 83 % similarity over a 216 amino acid stretch.

A polynucleotide encoding the GPR2 polypeptide of the present invention may be isolated from human liver, heart and leukocytes. The polynucleotide encoding GPR2 was discovered in a cDNA library derived from human adrenal gland tumor. It is structurally related to the G protein-coupled receptor family. It contains an open reading frame encoding a protein of 393 amino acid residues. The protein exhibits the highest degree of homology to human EDG-1 with 30 % identity and 52 % similarity over a 383 amino acid stretch. Potential ligands to GPR2 include but are not limited to anandamide, serotonin, adrenalin and noradrenalin.

A polynucleotide encoding the GPR3 polypeptide of the present invention may be isolated from human liver, kidney and pancreas. The polynucleotide encoding GPR3 was discovered in a cDNA library derived from human neutrophil. It is structurally related to the G protein-coupled receptor family. It contains an open reading frame encoding a protein of 293 amino acid residues. The protein exhibits the highest degree of homology to a human G-Protein Coupled Receptor open reading frame with 39 % identity and 61 % similarity over the entire amino acid sequence. Potential

ligands to GPR3 include but are not limited to platelet activating factor, thrombin, C5a and bradykinin.

A polynucleotide encoding the GPR4 polypeptide of the present invention may be found in human heart, spleen and leukocytes. The polynucleotide encoding GPR4 was discovered in a cDNA library derived from human twelve-week-old embryo. It is structurally related to the G-protein coupled receptor family. It contains an open reading frame encoding a protein of 344 amino acid residues. The protein exhibits the highest degree of homology to a chick orphan G-protein coupled receptor with 82 % identity and 91 % similarity over a 291 amino acid stretch. Potential ligands to GPR4 include but are not limited to thrombin, chemokine, and platelet activating factor.

The polynucleotides of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the mature polypeptides may be identical to the coding sequence shown in Figures 1-4 (SEQ ID No. 1, 3, 5 and 7) or that of the deposited clones or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same mature polypeptides as the DNA of Figures 1-4 (SEQ ID No. 1, 3, 5 and 7) or the deposited cDNAs.

The polynucleotides which encode for the mature polypeptides of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or for the mature polypeptides encoded by the deposited cDNAs may include: only the coding sequence for the mature polypeptide; the coding sequence for the mature polypeptide (and optionally additional coding sequence) and non-coding sequence, such as introns or non-coding sequence 5' and/or 3' of the coding sequence for the mature polypeptide.

Thus, the term "polynucleotid encoding a polypeptide" encompasses a polynucleotide which includes only coding s qu nce for th polypeptide as w ll as a polynucl otid which includes additional coding and/or non-coding sequence.

The present invention further relates to variants of the hereinabove described polynucleotides which encode for fragments, analogs and derivatives of the polypeptides having the deduced amino acid sequence of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or the polypeptides encoded by the cDNAs of the deposited clones. The variants of the polynucleotides may be a naturally occurring allelic variant of the polynucleotides or a non-naturally occurring variant of the polynucleotides.

Thus, the present invention includes polynucleotides encoding the same mature polypeptides as shown in Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or the same mature polypeptides encoded by the cDNAs of the deposited clones as well as variants of such polynucleotides which variants encode for a fragment, derivative or analog of the polypeptides of Figure 1-4 (SEQ ID No. 2, 4, 6 and 8) or the polypeptides encoded by the cDNAs of the deposited clones. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

As hereinabove indicated, the polynucleotides may have a coding sequence which is a naturally occurring allelic variant of the coding sequences shown in Figures 1-4 (SEQ ID No. 1, 3, 5 and 7) or of the coding sequences of the deposited clones. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptides.

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence which allows for purification of the polypeptides of the present invention. The marker sequence may be a hexa-

histidine tag supplied by a pQE-9 vector to provid for purification of the matur polypeptides fused to the marker in the case of a bacterial host, or, for example, the marker sequence may be a hemagglutinin (HA) tag when a mammalian host, e.g. COS-7 cells, is used. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein (Wilson, I., et al., Cell, 37:767 (1984)).

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 50% and preferably 70% identity between the sequences. The present invention particularly relates to polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which either retain substantially the same biological function or activity as the mature polypeptide encoded by the cDNAs of Figures 1-4 (SEQ ID No. 1, 3, 5 and 7) or the deposited cDNAs, i.e. function as a G-protein coupled receptor or retain the ability to bind the ligand for the receptor even though the polypeptides do not function as a G-protein coupled receptor, for example, soluble form of the receptors.

Alternatively, the polynucleotide may be a polynucleotide which has at least 20 bases, preferably 30 bases, and more preferably at least 50 bases which hybridize to a polynucleotide of the present invention and which has an identity thereto, as hereinabove described, and which does not retain activity. Such polynucleotides may be employed as probes for the polynucleotide of SEQ ID No. 1, for example, for recovery of the polynucleotide or as a diagnostic probe or as a PCR primer.

The deposit(s) referred to herein will be maintained under the terms of th Budapest Treaty on the Internati nal Recognition of the Dep sit of Micro-organisms for purposes f Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited materials, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

The present invention further relates to G-protein coupled receptor polypeptides which have the deduced amino acid sequences of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or which have the amino acid sequences encoded by the deposited cDNAs, as well as fragments, analogs and derivatives of such polypeptides.

The terms "fragment," "derivative" and "analog" when referring to the polypeptides of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or that encoded by the deposited cDNAs, means a polypeptide which either retains substantially the same biological function or activity as such polypeptide, i.e. functions as a G-protein coupled receptor, or retains the ability to bind the ligand or the receptor even though the polypeptide does not function as a G-protein coupled receptor, for example, a soluble form of the receptor. An analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an activ mature polypeptide.

The polypeptides of the present invention may be recombinant polypeptides, a natural polypeptides or synthetic polypeptides, preferably recombinant polypeptides.

Th fragment, derivative or analog of the polypeptides of Figures 1-4 (SEQ ID No. 2, 4, 6 and 8) or that encoded by the deposited cDNAs may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide which is employed for purification of the mature polypeptide. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the

invention and the production of polypeptides of the invention by r combinant techniques.

Host cells are genetically engin ered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the G-protein coupled receptor genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The polynucleotides of the present invention may be employed for producing polypeptides by recombinant techniques. Thus, for example, the polynucleotide may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative

examples of such promoters, there may be mentioned: LTR or SV40 promoter, the E. coli. lac or trp, th phag lambda P_L promoter and other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in E. coli.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

As representative examples of appropriate hosts, there may be mentioned: bacterial cells, such as E. coli, Streptomyces, Salmonella typhimurium; fungal cells, such as yeast; insect cells such as Drosophila S2 and Spodoptera Sf9; animal cells such as CHO, COS or Bowes melanoma; adenoviruses; plant cells, etc. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors

and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); pTRC99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, et al., Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act on a promoter to increase its transcription. Examples including the SV40 enhancer on the late side of the replication origin bp 100 to 270, a cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of E. coli and S. cerevisiae TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. Such promoters can be derived from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), α-factor, acid phosphatase, or heat shock proteins, among others. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA sequence encoding a desired protein together with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host. Suitable prokaryotic hosts for transformation include E. coli, Bacillus subtilis, Salmonella typhimurium and various species within the genera Pseudomonas, Streptomyces, and Staphylococcus, although others may also be employed as a matter of choice.

As a representative but nonlimiting example, useful expression vectors for bacterial use can comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and GEM1 (Promega Biotec, Madison, WI, USA). These pBR322 "backbone" sections are combined with an appropriate promoter and the structural sequence to be expressed.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell

lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts, described by Gluzman, Cell, 23:175 (1981), and other cell lines capable of expressing a compatible vector, for example, the C127, 3T3, CHO, HeLa and BHK cell lines. Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. DNA sequences derived from the SV40 splice, and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

The G-protein coupled receptor polypeptides can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

The polypeptides of the present invention may be a naturally purified product, or a product of chemical synthetic procedures, or produced by recombinant techniques from a prokaryotic or eukaryotic host (for example, by bacterial, yeast, higher plant, insect and mammalian cells in culture). Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated.

Polypeptides of the invention may also include an initial methionine amino acid residue.

Fragments of the full length G-protein coupled receptor genes may be employed as a hybridization probe for a cDNA library to isolate the full length genes and to isolate other genes which have a high sequence similarity to the gene or similar biological activity. Probes of this type generally have at least 20 bases. Preferably, however, the probes have at least 30 bases and may contain, for example, 50 bases or more. In many cases, the probe has from 20 to 50 bases. The probe may also be used to identify a cDNA clone corresponding to a full length transcript and a genomic clone or clones that contain the complete G-protein coupled receptor gene including regulatory and promotor regions, exons, and introns. As an example of a screen comprises isolating the coding region of the G-protein coupled receptor gene by using the known DNA sequence to synthesize an oligonucleotide probe. Labeled oligonucleotides having a sequence complementary to that of the gene of the present invention are used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The G-protein coupled receptors of the present invention may be employed in a process for screening for antagonists and/or agonists for the receptor.

In general, such screening procedures involve providing appropriate cells which express the receptor on the surface thereof. Such cells include cells from mammals, yeast, drosophila or *E. Coli*. In particular, a polynucleotide encoding the receptor of the present invention is employed to transfect cells to thereby express the respective G-protein coupled receptor. The expressed receptor is then contacted with a test compound to observe binding, stimulation or inhibition of a functional response.

One such screening procedure involves the use of melanophores which are transfected to express the respective G-protein coupled receptor of the present invention. Such a screening technique is described in PCT WO 92/01810 published February 6, 1992.

Thus, for example, such assay may be employed for screening for a receptor antagonist by contacting the melanophore cells which encode the G-protein coupled receptor with both the receptor ligand and a compound to be screened. Inhibition of the signal generated by the ligand indicates that a compound is a potential antagonist for the receptor, i.e., inhibits activation of the receptor.

The screen may be employed for determining an agonist by contacting such cells with compounds to be screened and determining whether such compound generates a signal, i.e., activates the receptor.

Other screening techniques include the use of cells which express the G-protein coupled receptor (for example, transfected CHO cells) in a system which measures extracellular pH changes caused by receptor activation, for example, as described in Science, volume 246, pages 181-296 (October 1989). For example, potential agonists or antagonists may be contacted with a cell which expresses the G-protein coupled receptor and a second messenger response, e.g. signal transduction or pH changes, may be measured to determine whether the potential agonist or antagonist is effective.

Another such screening technique involves introducing RNA encoding the G-protein coupled receptors into *Xenopus* oocytes to transiently express the receptor. The receptor oocytes may then be contacted in the case of antagonist screening with the receptor ligand and a compound to be screened, followed by detection of inhibition of a calcium signal.

Another screening technique involves expressing the G-protein coupled receptors in which the receptor is linked to a phosph lipase C or D. As representative examples of such cells, there may be mentioned endothelial cells, smooth muscle cells, embryonic kidney cells, etc. The screening for an antagonist or agonist may be accomplished as hereinabove described by detecting activation of the receptor or inhibition of activation of the receptor from the phospholipase second signal.

Another method involves screening for antagonists by determining inhibition of binding of labeled ligand to cells which have the receptor on the surface thereof. Such a method involves transfecting a eukaryotic cell with DNA encoding the G-protein coupled receptor such that the cell expresses the receptor on its surface and contacting the cell with a potential antagonist in the presence of a labeled form of a known ligand. The ligand can be labeled, e.g., by radioactivity. The amount of labeled ligand bound to the receptors is measured, e.g., by measuring radioactivity of the receptors. If the potential antagonist binds to the receptor as determined by a reduction of labeled ligand which binds to the receptors, the binding of labeled ligand to the receptor is inhibited.

G-protein coupled receptors are ubiquitous in the mammalian host and are responsible for many biological functions, including many pathologies. Accordingly, it is desirous to find compounds and drugs which stimulate the G-protein coupled receptors on the one hand and which can antagonize a G-protein coupled receptor on the other hand, when it is desirable to inhibit the G-protein coupled receptor.

For example, agonists for G-protein coupled receptors may be employed for therapeutic purposes, such as the treatment of asthma, Parkinson's disease, acute heart failure, hypotension, urinary retention, and osteoporosis.

In general, antagonists to the G-protein coupled receptors may be employed for a variety of therapeutic purposes, for example, for the treatment of hypertension, angina pectoris, myocardial infarction, ulcers, asthma, allergies, benign prostatic hypertrophy and psychotic and neurological disorders, including schizophrenia, manic excitement, depression, delirium, dementia or severe mental retardation, dyskinesias, such as Huntington's disease or Gilles dila Tourett's syndrome, among others. G-protein coupled receptor antagonists have also been useful in reversing endogenous anorexia and in the control of bulimia.

Examples of G-protein coupled receptor antagonists include an antibody, or in some cases an oligopeptide, which binds to the G-protein coupled receptors but does not elicit a second messenger response such that the activity of the G-protein coupled receptors is prevented. Antibodies include anti-idiotypic antibodies which recognize unique determinants generally associated with the antigen-binding site of an antibody. Potential antagonists also include proteins which are closely related to the ligand of the G-protein coupled receptors, i.e. a fragment of the ligand, which have lost biological function and when binding to the G-protein coupled receptors, elicit no response.

A potential antagonist also includes an antisense construct prepared through the use of antisense technology. Antisense technology can be used to control gene expression through triple-helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide to DNA or RNA. For example, the 5' coding portion of the polynucleotide sequence, which encodes for the mature polypeptides of the present invention, is used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription (triple helix -see Lee et al., Nucl. Acids

Res., 6:3073 (1979); Cooney et al, Science, 241:456 (1988); and Dervan et al., Science, 251: 1360 (1991)), thereby preventing transcription and the production of G-protein coupled receptors. The antisense RNA oligonucleotide hybridizes to the mRNA *in vivo* and blocks translation of mRNA molecules into G-protein coupled receptors (antisense - Okano, J. Neurochem., 56:560 (1991); Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988)). The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed *in vivo* to inhibit production of G-protein coupled receptors.

Another potential antagonist is a small molecule which binds to the G-protein coupled receptor, making it inaccessible to ligands such that normal biological activity is prevented. Examples of small molecules include but are not limited to small peptides or peptide-like molecules.

Potential antagonists also include a soluble form of a G-protein coupled receptor, e.g. a fragment of the receptors, which binds to the ligand and prevents the ligand from interacting with membrane bound G-protein coupled receptors.

This invention additionally provides a method of treating an abnormal condition related to an excess of G-protein coupled receptor activity which comprises administering to a subject the antagonist as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to block binding of ligands to the G-protein coupled receptors and thereby alleviate the abnormal conditions.

The invention also provides a method of treating abnormal conditions related to an under-expression of G-protein coupled receptor activity which comprises administering to a subject a therapeutically effective amount of the agonist described above in combination with a pharmaceutically acceptable carrier, in an amount effective

to nhance binding of ligands to the G-protein coupl d receptor and thereby alleviate the abnormal conditions.

The soluble form of the G-protein coupl d receptors, antagonists and agonists may be employed in combination with a suitable pharmaceutical carrier. Such compositions comprise a therapeutically effective amount of the antagonist or agonist, and a pharmaceutically acceptable carrier or excipient. Such a carrier includes but is not limited to saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The formulation should suit the mode of administration.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the pharmaceutical compositions may be employed in conjunction with other therapeutic compounds.

The pharmaceutical compositions may be administered in a convenient manner such as by the topical, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes. The pharmaceutical compositions are administered in an amount which is effective for treating and/or prophylaxis of the specific indication. In general, the pharmaceutical compositions will be administered in an amount of at least about 10 μg/kg body weight and in most cases they will be administered in an amount not in excess of about 8 mg/Kg body weight per day. In most cases, the dosage is from about 10 μg/kg to about 1 mg/kg body weight daily, taking into account the routes of administration, symptoms, etc.

The G-protein coupled receptor polypeptides, and antagonists or agonists which are polypeptides, may be employed in accordance with the present invention by expression of such polypeptides *in vivo*, which is often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide (DNA or RNA) encoding a polypeptide *ex vivo*, with the engineered cells then being provided to a patient to be treated with the polypeptide. Such methods are well-known in the art. For example, cells may be engineered by procedures known in the art by use of a retroviral particle containing RNA encoding a polypeptide of the present invention.

Similarly, cells may be engineered *in vivo* for expression of a polypeptide *in vivo* by, for example, procedures known in the art. As known in the art, a producer cell for producing a retroviral particle containing RNA encoding the polypeptide of the present invention may be administered to a patient for engineering cells *in vivo* and expression of the polypeptide *in vivo*. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention. For example, the expression vehicle for engineering cells may be other than a retrovirus, for example, an adenovirus which may be used to engineer cells in vivo after combination with a suitable delivery vehicle.

The present invention also provides a method for determining whether a ligand not known to be capable of binding to a G-protein coupled receptor can bind to such receptor which comprises contacting a mammalian cell which expresses a G-protein coupled receptor with the ligand under conditions permitting binding of ligands to the G-protein coupled receptor, detecting the presence of a ligand which

binds to the rec ptor and thereby determining whether the ligand binds to the G-protein c upled receptor.

This invention further provides a method of screening drugs to identify drugs which specifically interact with, and bind to, the human G-protein coupled receptors on the surface of a cell which comprises contacting a mammalian cell comprising an isolated DNA molecule encoding the G-protein coupled receptor with a plurality of drugs, determining those drugs which bind to the mammalian cell, and thereby identifying drugs which specifically interact with and bind to a human G-protein coupled receptor of the present invention.

This invention also provides a method of detecting expression of the G-protein coupled receptor on the surface of a cell by detecting the presence of mRNA coding for a G-protein coupled receptor which comprises obtaining total mRNA from the cell and contacting the mRNA so obtained with a nucleic acid probe comprising a nucleic acid molecule of at least 15 nucleotides capable of specifically hybridizing with a sequence included within the sequence of a nucleic acid molecule encoding a human G-protein coupled receptor under hybridizing conditions, detecting the presence of mRNA hybridized to the probe, and thereby detecting the expression of the G-protein coupled receptor by the cell.

This invention is also related to the use of the G-protein coupled receptor genes as part of a diagnostic assay for detecting diseases or susceptibility to diseases related to the presence of mutations in the G-protein coupled receptor genes. Such diseases, by way of example, are related to cell transformation, such as tumors and cancers.

Individuals carrying mutations in the human G-protein coupled receptor genes may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA

may be used directly for detection or may be amplified enzymatically by using PCR (Saiki *et al.*, Nature, 324:163-166 (1986)) prior to analysis. RNA or cDNA may also be used for the same purpose. As an example, PCR primers complementary to the nucleic acid encoding the G-protein coupled receptor proteins can be used to identify and analyze G-protein coupled receptor mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled G-protein coupled receptor RNA or alternatively, radiolabeled G-protein coupled receptor antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Sequence differences between the reference gene and genes having mutations may be revealed by the direct DNA sequencing method. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial

melting temperatures (see, e.g., Myers *et al.*, Science, 230:1242 (1985)).

Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton *et al.*, PNAS, USA, 85:4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., Restriction Fragment Length Polymorphisms (RFLP)) and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations can also be detected by *in situ* analysis.

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence *in situ* hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clones from which the EST was derived, and the longer the better. For example, 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the

affected individuals but not in any normal individuals, then the mutation is likely to b the causative agent of the diseas .

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies

(Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

The present invention will be further described with reference to the following examples; however, it is to be understood that the present invention is not limited to such examples. All parts or amounts, unless otherwise specified, are by weight.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 μg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 μl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in a larger

volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, Nucleic Acids Res., 8:4057 (1980).

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 μg of approximately equimolar amounts of the DNA fragments to be ligated.

Unless otherwise stated, transformation was performed as described in the method of Graham, F. and Van der Eb, A., Virology, 52:456-457 (1973).

Example 1

Bacterial Expression and Purification of GPR1

The DNA sequence encoding GPR1, ATCC # 75981, is initially amplified using PCR oligonucleotide primers corresponding to the 5' and 3' end sequences of the processed G-protein coupled receptor nucleotide sequence. Additional

nucleotid s corresponding to the GPR1 nucleotide sequence are added to the 5' and 3' sequences respectiv ly. Th 5' lig nucleotide primer has the sequenc 5' GACTAAAGCTTATGAGTAGTGAAATGGTG 3' (SEQ ID No. 9) contains a HindIII restriction enzyme site followed by 19 nucleotides of G-protein coupled receptor coding sequence starting from the presumed terminal amino acid of the processed protein. The 3' sequence 5' GAACTTCTAGACCCTCAGGGTTGTAAATCAG 3' (SEQ ID No. 10) contains complementary sequences to an XbaI site and is followed by 20 nucleotides of GPR1 coding sequence. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-9 (Qiagen, Inc. Chatsworth, CA). pQE-9 encodes antibiotic resistance (Amp'), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-9 is then digested with HindIII and XbaI. The amplified sequences are ligated into pQE-9 and are inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform E. coli strain M15/rep 4 (Qiagen, Inc.) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan'). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ($O.D._{600}$) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-

thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating th lacI repressor, clearing the P/O leading to increased gene expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized G-protein coupled receptor is purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). GPR1 is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

Example 2

Bacterial Expression and Purification of GPR2

The DNA sequence encoding GPR2, ATCC # 75983, is initially amplified using PCR oligonucleotide primers corresponding to the 5' and 3' end sequences of the processed GPR2 coding sequence. Additional nucleotides corresponding to GPR2 coding sequence are added to the 5' and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' GACTAAAGCTTAATGAGGCCCACATGGGCA 3' (SEQ ID No. 11) contains a HindIII restriction enzyme site followed by 19 nucleotides of GPR2 coding sequence starting from the presumed terminal amino acid of the processed protein. The 3' sequence 5' GAACTTCTAGACGAACTAGTGGATCCCCCCGG 3' (SEQ ID No. 12) contains complementary sequences to an XbaI site and is followed by 21 nucleotides of GPR2 coding sequence. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-9 (Qiagen, Inc. Chatsworth,

CA). pQE-9 encodes antibiotic resistanc (Amp^r), a bacterial origin of replication (ori), an IPTG-r gulatabl promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-9 is then digested with HindIII and XbaI. The amplified sequences are ligated into pQE-9 and are inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform E. coli strain M15/rep 4 (Qiagen, Inc.) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.^{600}) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized GPR2 is purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). GPR2 is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium

phosphate, 10 mmolar glutathione (reduc d) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

Example 3

Bacterial Expression and Purification of GPR3

The DNA sequence encoding GPR3, ATCC # 75976, is initially amplified using PCR oligonucleotide primers corresponding to the 5' and 3' end sequences of the processed G-protein coupled receptor nucleotide sequence. Additional nucleotides corresponding to the GPR3 coding sequence are added to the 5' and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' GACTAAAGCTTAATGGCGTCTTTCTCTGCT 3' (SEQ ID No. 13) contains a HindIII restriction enzyme site followed by 19 nucleotides of GPR3 coding sequence starting from the presumed terminal amino acid of the processed protein. The 3' sequence 5' GAACTTCTAGACTTCACACAGTTGTACTAT 3' (SEQ ID No. 14) contains complementary sequences to XbaI site and is followed by 19 nucleotides of GPR3 coding sequence. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-9 (Qiagen, Inc. Chatsworth, CA). pQE-9 encodes antibiotic resistance (Amp'), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-9 is then digested with XbaI and XbaI. The amplified sequences are ligated into pQE-9 and are inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform E. coli strain M15/rep 4 (Qiagen Inc.) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers

kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies ar selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.^{600}) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized GPR3 is purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). GPR3 is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

Example 4

Bacterial Expression and Purification of GPR4

The DNA sequence encoding GPR4, ATCC # 75979, is initially amplified using PCR oligonucleotide primers corresponding to the 5' and 3' end sequences of the processed GPR4 nucleotide sequence. Additional nucleotides corresponding to the GPR4 coding sequence are added to the 5'

and 3' sequences respectively. The 5' oligonucleotide primer has the sequence 5' GACTAAAGCTTAATGGTAAGCGTTAACAGC 3' (SEQ ID No. 15) contains a HindIII restriction enzyme site followed by 19 nucleotides of GPR4 coding sequence starting from the presumed terminal amino acid of the processed protein. The 3' sequence 5' GAACTTCTAGACTTCAGGCAGCAGATTCATT 3' (SEQ ID No. 16) contains complementary sequences to XbaI site and is followed by 20 nucleotides of GPR4 coding sequence. The restriction enzyme sites correspond to the restriction enzyme sites on the bacterial expression vector pQE-9 (Qiagen, Inc. Chatsworth, CA). pQE-9 encodes antibiotic resistance (Amp^r), a bacterial origin of replication (ori), an IPTG-regulatable promoter operator (P/O), a ribosome binding site (RBS), a 6-His tag and restriction enzyme sites. pQE-9 is then digested with HindIII and XbaI. The amplified sequences are ligated into pQE-9 and are inserted in frame with the sequence encoding for the histidine tag and the RBS. The ligation mixture is then used to transform E. coli strain M15/rep 4 (Qiagen, Inc.) by the procedure described in Sambrook, J. et al., Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989). M15/rep4 contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan^r). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.^{600}) of between 0.4 and 0.6. IPTG ("Isopropyl-B-D-thiogalacto pyranoside") is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene

expression. Cells are grown an extra 3 to 4 hours. Cells are then harvested by centrifugation. The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized GPR4 is purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). GPR4 is eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein is dialyzed to 10 mmolar sodium phosphate.

Example 5

Expression of Recombinant GPR1 in COS cells

The expression of plasmid, GPR1 HA is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, a SV40 intron and polyadenylation site. A DNA fragment encoding the entire GPR1 precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag correspond to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37, 767). The infusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding GPR1, ATCC # 75981, is constructed by PCR using two primers: the 5' primer 5' GTCCAAGCTTGCCACCATGAGTAGTGAAATGGTG 3' (SEQ ID No. 17) contains a HindIII site followed by 18 nucleotides of GPR1 coding sequence starting from the initiation codon; the 3' sequence 5' CTAGCTCGAGTCAAGCGTAGTCTGGGACGTCGTATGGGTAGCAGG GTTGTAAATCAGG 3' (SEQ ID No. 18) contains complementary sequences to an XhoI site, translation stop codon, HA tag and the last 15 nucleotides of the GPR1 coding sequence (not including the stop codon). Therefore, the PCR product contains a HindIII site, GPR1 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XhoI site. The PCR amplified DNA fragment and the vector, pcDNAI/Amp, are digested with HindIII and XhoI restriction enzymes and ligated. The ligation mixture is transformed into E. coli strain SURE (Stratagene Cloning Systems, La Jolla, CA) the transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant GPR1, COS cells are transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the GPR1 HA protein is detected by radiolabeling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells are labelled for 8 hours with ^{35}S-cysteine two days post transfection. Culture media are then collected and cells are lysed with detergent (RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5). (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media are precipitated with a HA specific monoclonal

antibody. Proteins precipitated are analyzed on 15% SDS-PAGE gels.

Example 6

Expression of Recombinant GPR2 in COS cells

The expression of plasmid, GPR2 HA is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, a SV40 intron and polyadenylation site. A DNA fragment encoding the entire GPR2 precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag correspond to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37, 767). The infusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding for GPR2, ATCC # 75983, is constructed by PCR using two primers: the 5' primer 5' GTCCAAGCTTGCCACCATGGTTGGTGGCACCTGG 3' (SEQ ID No. 19) contains an HindIII site followed by 18 nucleotides of GPR2 coding sequence starting from the initiation codon; the 3' sequence 5' CTAGCTCGAGTCAAGCGTAGTCTGGGACGTCGTATGGGTAGCAGTG GATCCCCCGTGC 3' (SEQ ID No. 20) contains complementary sequences to an XhoI site, translation stop codon, HA tag and the last 15 nucleotides of the GPR2 coding sequence (not including the stop codon). Therefore, the PCR product contains a HindIII site, GPR2 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XhoI site. The PCR amplified DNA

fragment and the vector, pcDNAI/Amp, are digested with HindIII and XhoI restriction enzymes and ligated. The ligation mixtur is transformed into E. coli strain SURE (Stratagene Cloning Systems, La Jolla, CA) the transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant GPR2, COS cells are transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the GPR2 HA protein is detected by radiolabelling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells are labelled for 8 hours with ^{35}S-cysteine two days post transfection. Culture media are then collected and cells are lysed with detergent (RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5). (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media are precipitated with a HA specific monoclonal antibody. Proteins precipitated are analyzed on 15% SDS-PAGE gels.

Example 7

Expression of Recombinant GPR3 in COS cells

The expression of plasmid, GPR3 HA is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, a SV40 intron and polyadenylation site. A DNA fragment encoding the entire GPR3 precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag

correspond to an epitop derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37, 767). The infusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding for GPR3, ATCC # 75976, is constructed by PCR using two primers: the 5' primer 5' GTCCAAGCTTGCCACCATGAACACCACAGTAATG 3' (SEQ ID No. 21) contains a HindIII site followed by 18 nucleotides of GPR3 coding sequence starting from the initiation codon; the 3' sequence 5' CTAGCTCGAGTCAAGCGTAGTCTGGGACGTCGTATGGGTAGCAAGG GATCCATACAAATGT 3' (SEQ ID No. 22) contains complementary sequences to an XhoI site, translation stop codon, HA tag and the last 18 nucleotides of the GPR3 coding sequence (not including the stop codon). Therefore, the PCR product contains a HindIII site, GPR3 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XhoI site. The PCR amplified DNA fragment and the vector, pcDNAI/Amp, are digested with HindIII and XhoI restriction enzymes and ligated. The ligation mixture is transformed into E. coli strain SURE (Stratagene Cloning Systems, La Jolla, CA) the transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant GPR3, COS cells are transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the GPR3 HA protein is detected by radiolabelling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies:

A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells are labell d for 8 hours with ^{35}S-cysteine two days post transfection. Culture media are then collected and cells are lysed with detergent (RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5) (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media are precipitated with a HA specific monoclonal antibody. Proteins precipitated are analyzed on 15% SDS-PAGE gels.

Example 8

Expression of Recombinant GPR4 in COS cells

The expression of plasmid, GPR4 HA is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, a SV40 intron and polyadenylation site. A DNA fragment encoding the entire GPR4 precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag correspond to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37, 767). The infusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding for GPR4, ATCC # 75979, is constructed by PCR using two primers: the 5' primer 5' GTCCAAGCTTGCCACCATGGTAAGCGTTAACAGC 3' (SEQ ID No. 23) contains a HindIII site followed by 18 nucleotides of GPR4 coding sequence starting from the initiation codon; the 3' sequence 5' CTAGCTCGAGTCAAGCGTAGTCTGGGACGTCGTATGGGTAGCAGG

CAGCAGATTCATTGTC 3' (SEQ ID No. 24) contains complementary sequences to an XhoI site, translation stop codon, HA tag and the last 18 nucleotides of the GPR4 coding sequence (not including the stop codon). Therefore, the PCR product contains a HindIII site, GPR4 coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and an XhoI site. The PCR amplified DNA fragment and the vector, pcDNAI/Amp, are digested with Hind III and XhoI restriction enzymes and ligated. The ligation mixture is transformed into E. coli strain SURE (Stratagene Cloning Systems, La Jolla, CA) the transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant GPR4, COS cells are transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the GPR4 HA protein is detected by radiolabelling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells are labelled for 8 hours with ^{35}S-cysteine two days post transfection. Culture media are then collected and cells are lysed with detergent (RIPA buffer (150 mM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5) (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media are precipitated with a HA specific monoclonal antibody. Proteins precipitated are analyzed on 15% SDS-PAGE gels.

Example 9

Cloning and expression of GPR1 using the baculovirus expression system

The DNA sequence encoding the full length GPR1 protein, ATCC # 75981, is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of th gene:

The 5' primer has the sequence 5' CGGGATCCCTCCATGAG TAGTGAAATGGTG 3' (SEQ ID No. 25) and contains a BamHI restriction enzyme site (in bold) followed by 4 nucleotides resembling an efficient signal for the initiation of translation in eukaryotic cells (Kozak, M., J. Mol. Biol., 196:947-950 (1987) which is just behind the first 18 nucleotides of the GPR1 gene (the initiation codon for translation "ATG" is underlined).

The 3' primer has the sequence 5' CGGGATCCCGCT CAGGGTTGTAAATCAGG 3' (SEQ ID No. 26) and contains the cleavage site for the BamHI restriction endonuclease and 18 nucleotides complementary to the 3' non-translated sequence of the GPR1 gene. The amplified sequences are isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment is then digested with the endonuclease BamHI and then purified again on a 1% agarose gel. This fragment is designated F2.

The vector pRG1 (modification of pVL941 vector, discussed below) is used for the expression of the GPR1 protein using the baculovirus expression system (for review see: Summers, M.D. and Smith, G.E. 1987, A manual of methods for baculovirus vectors and insect cell culture procedures, Texas Agricultural Experimental Station Bulletin No. 1555). This expression vector contains the strong polyhedrin promoter of the Autographa californica nuclear polyhedrosis virus (AcMNPV) followed by the recognition sites for the restriction endonuclease BamHI. The polyadenylation site of the simian virus (SV)40 is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from E.coli is inserted in the same orientation as the polyhedrin promoter followed by the polyadenylation signal of the polyhedrin gene. The

polyhedrin sequences are flanked at both sides by viral sequences for th cell-mediated homologous recombination of cotransfected wild-type viral DNA. Many other baculovirus vectors could be used in place of pRG1 such as pAc373, pVL941 and pAcIM1 (Luckow, V.A. and Summers, M.D., Virology, 170:31-39).

The plasmid is digested with the restriction enzymes BamHI and then dephosphorylated using calf intestinal phosphatase by procedures known in the art. The DNA is then isolated from a 1% agarose gel using the commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated V2.

Fragment F2 and the dephosphorylated plasmid V2 are ligated with T4 DNA ligase. E.coli HB101 cells are then transformed and bacteria identified that contained the plasmid (pBacGPR1) with the GPR1 gene using the enzymes BamHI. The sequence of the cloned fragment is confirmed by DNA sequencing.

5 μg of the plasmid pBacGPR1 is cotransfected with 1.0 μg of a commercially available linearized baculovirus ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA.) using the lipofection method (Felgner et al. Proc. Natl. Acad. Sci. USA, 84:7413-7417 (1987)).

1μg of BaculoGold™ virus DNA and 5 μg of the plasmid pBacGPR1 are mixed in a sterile well of a microtiter plate containing 50 μl of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10 μl Lipofectin plus 90 μl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added dropwise to the Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is rocked back and forth to mix the newly added solution. The plate is then incubated for 5 hours at 27°C. After 5 hours the transfection solution is removed from the plate and 1 ml of

Grace's insect medium supplemented with 10% fetal calf serum is added. The plate is put back into an incubator and cultivation continued at 27°C for four days.

After four days the supernatant is collected and a plaque assay performed similar as described by Summers and Smith (supra). As a modification an agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used which allows an easy isolation of blue stained plaques. (A detailed description of a "plaque assay" can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after the serial dilution, the virus are added to the cells, blue stained plaques are picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses is then resuspended in an Eppendorf tube containing 200 µl of Grace's medium. The agar is removed by a brief centrifugation and the supernatant containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then stored at 4°C.

Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus V-GPR1 at a multiplicity of infection (MOI) of 2. Six hours later the medium is removed and replaced with SF900 II medium minus methionine and cysteine (Life Technologies Inc., Gaithersburg). 42 hours later 5 µCi of ^{35}S-methionine and 5 µCi ^{35}S cysteine (Amersham) are added. The cells are further incubated for 16 hours before they are harvested by centrifugation and the labelled proteins visualized by SDS-PAGE and autoradiography.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, within the scope of the appended claims, the

invention may be practiced otherwise than as particularly described.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: LI, ET AL.

(ii) TITLE OF INVENTION: Human G-Protein Coupled Receptors

(iii) NUMBER OF SEQUENCES: 26

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEIN
(B) STREET: 6 BECKER FARM ROAD
(C) CITY: ROSELAND
(D) STATE: NEW JERSEY
(E) COUNTRY: USA
(F) ZIP: 07068

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: 3.5 INCH DISKETTE
(B) COMPUTER: IBM PS/2
(C) OPERATING SYSTEM: MS-DOS
(D) SOFTWARE: WORD PERFECT 5.1

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: Concurrently
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA
(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: FERRARO, GREGORY D.
(B) REGISTRATION NUMBER: 36,134
(C) REFERENCE/DOCKET NUMBER: 325800-270

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 201-994-1700
(B) TELEFAX: 201-994-1744

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 1713 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
GGCACGAGGT CATTCAACAT TTATTCAACC AAAAATACTA AGTCAGCTCT ATACAAACTA    60
ATGGAAGGAT ACAGCTATGC AAATATAGAA CACTAAAGTG TTACATGACA GATGTATGAG   120
TAGTGAAATG GTGAAAAATC AGACAATGGT CACAGAGTTC CTCCTACTGG GATTTCTCCT   180
GGGCCCAAGG ATTCAGATGC TCCTCTTTGG GCTCTTCTCC CTGTTCTATG TCTTCACCCT   240
GCTGGGGAAT GGGACCATCC TGGGGCTCAT CTCACTGGAC TCCAGACTCC ACACCCCCAT   300
GTACTTCTTC CTCTCACACC TGGCCGTCGT CAACATCGCC TATGCCTGCA ACACAGTGCC   360
CCAGATGCTG GTGAACCTCC TGCATCCAGC CAAGCCCATC TCCTTTGCTG GTTGCATGAC   420
ACTAGACTTT CTCTTTTTGA GTTTTGCACA TACTGAATGC CTCCTGTTGG TGCTGATGTC   480
CTACGATCGG TACGTGGCCA TCTGCCACCC TCTCCGATAT TTCATCATCA TGACCTGGAA   540
AGTCTGCATC ACTCTGGGCA TCACTTCCTG GACATGTGGC TCCCTCCTGG CTATGGTCCA   600
TGTGAGCCTC ATCCTAAGAC TGCCCTTTTG TGGGCCTCGT GAAATCAACC ACTTCTTCTG   660
TGAAATCCTG TCTGTCCTCA GGCTGGCCTG TGCTGATACC TGGCTCAACC AGGTGGTCAT   720
CTTTGAAGCC TGCATGTTCA TCCTGGTGGG ACCACTCTGC CTGGTGCTGG TCTCCTACTC   780
ACACATCCTG GGGGGCATCC TGAGGATCCA GTCTGGGGAG GGCCGCAGAA AGGCCTTCTC   840
CACCTGCTCC TCCCACCTCT GCGTAGTGGG ACTCTTCTTT GGSAGCGCCA TCGTCATGTA   900
CATGGCCCCT AAGTCCCGCC ATCCTGAGGA GCAGCAGAAG GTCCTTTTTC TTATTTTACA   960
GTTCCTTTCA ACCCCGATGC TTAAACCCCC TGATTTACAA CCCTGAGGAA TGTAGAGGGT  1020
CAAGGGTGCC CTCCGAGGAG ACCACTGTGC AARGRAAGTC ATTCCTAAGG GGTGTGACAT  1080
TTGAACTGCC AGCCCCAGTT GCCCCGTGGA CTCCTGATGC CCAATTATTG CCTCAACCCA  1140
GAAAAGTTTA CTCCCCTTTA ACTGTGCTTT ACTGACAGAA GGGCAAGCCT TCTCCCGTTT  1200
TTTGCAGATA AAATTTTAGA TGTGTTGCAA TCATTGGGTT TCTAGGAGAT GTGGTTTTAT  1260
CAGACAATTT TTTCTTTTAT TTCACAATTA CTTTAATATC TGTAAAATAA AGAATTATTT  1320
TAAATCATTT TCCCAGTCCC AAAAGTTAAA TACAGGCCAC TTACTTCTTT AACCAAATGA  1380
TATAGTTTGG CTCTGTGTCC CCACCCAAAT CTCATGTCAA ATTGTAATCC CCGCATGTCA  1440
GCGGAGGGAC CTGGTGGGAG GTGATTGGAT CATGGGGAGG GATTTCCCCC TTGCTGTTCT  1500
GTTGATAGTG AACGAGTTCT CACGAAATCT GATGGTTTAA AAGTGCAGCA CTTCTCCCTT  1560
TGCTCTCTCT CTCCTGCTGT GCCATGGTAA GACGTGCCTT GCTTCCCCTG GTGCTTCCGC  1620
CATGATTGTA CCTTTCCTGA GGCCTCTCCA GCCATGTGGA ACTGTGAGCC AATTAAACTT  1680
CTTTTCTTTA GAAAAAAAAA AAAAAAAAAA AAA                               1713
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 296 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) STRANDEDNESS:

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ser Ser Glu Met Val Lys Asn Gln Thr Met Val Thr Glu Phe
 5                  10                  15
Leu Leu Leu Gly Phe Leu Leu Gly Pro Arg Ile Gln Met Leu Leu
            20                  25                  30
Phe Gly Leu Phe Ser Leu Phe Tyr Val Phe Thr Leu Leu Gly Asn
        35                  40                  45
Gly Thr Ile Leu Gly Leu Ile Ser Leu Asp Ser Arg Leu His Thr
    50                  55                  60
Pro Met Tyr Phe Phe Leu Ser His Leu Ala Val Val Asn Ile Ala
65                  70                  75
Tyr Ala Cys Asn Thr Val Pro Gln Met Leu Val Asn Leu Leu His
            80                  85                  90
Pro Ala Lys Pro Ile Ser Phe Ala Gly Cys Met Thr Leu Asp Phe
```

```
                    95                  100                 105
Leu Phe Leu Ser Phe Ala His Thr Glu Cys Leu Leu Leu Val L u
                   110                 115                 120
Met Ser Tyr Asp Arg Tyr Val Ala Ile Cys His Pro Leu Arg Tyr
                   125                 130                 135
Phe Ile Ile Met Thr Trp Lys Val Cys Ile Thr Leu Gly Ile Thr
                   140                 145                 150
Ser Trp Thr Cys Gly Ser Leu Leu Ala Met Val His Val Ser Leu
                   155                 160                 165
Ile Leu Arg Leu Pro Phe Cys Gly Pro Arg Glu Ile Asn His Phe
                   170                 175                 180
Phe Cys Glu Ile Leu Ser Val Leu Arg Leu Ala Cys Ala Asp Thr
                   185                 190                 195
Trp Leu Asn Gln Val Val Ile Phe Glu Ala Cys Met Phe Ile Leu
                   200                 205                 210
Val Gly Pro Leu Cys Leu Val Leu Val Ser Tyr Ser His Ile Leu
                   215                 220                 225
Gly Gly Ile Leu Arg Ile Gln Ser Gly Glu Gly Arg Arg Lys Ala
                   230                 235                 240
Phe Ser Thr Cys Ser Ser His Leu Cys Val Val Gly Leu Phe Phe
                   245                 250                 255
Gly Ser Ala Ile Val Met Tyr Met Ala Pro Lys Ser Arg His Pro
                   260                 265                 270
Glu Glu Gln Gln Lys Val Leu Phe Leu Ile Leu Gln Phe Leu Ser
                   275                 280                 285
Thr Pro Met Leu Lys Pro Pro Asp Leu Gln Pro
                   290                 295
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 2185 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TCACTATAGG GCGAATTGGG TACGGGCCCC CCCTCGAGGT CGACGGTATC GATAAGCTTG    60
ATATCGAATT CGGCACGAGC CGGGCTCGGA GAGGTGACGG AACCGGGGCT GGTAGCATAG   120
TTTGATTTGA TGATGGAGCC AACACAGGGG TTGGAGCTGG TACCGGTGAA GCTGAGGCTA   180
AAAAGGTTCC TGGAGTAGAC GATGGAGCCA TAACTGGAAC CGGAGTCTGT GAATGAAGCC   240
AGGACAGGAG CAGCACCTGG CGATGGTGCC AGGACCGGAA GAGGAGCCAG AGGAGGAGCT   300
GGAGAAGGAG CCAGAATTGC TGTCTGTGGA GCCGCCATAG GAGCCAGAGG GGTGGCTAGA   360
GCCTGAGAAT GCAGAAGATG CTGGAGCCAG AAGGGAAGCC TGAGCTGGAG CTGGATTTGG   420
TGCTGACGGA AAAGGACTGG CCAGAGCCGA AGCTGGCACC AGGGACAGGT GAGCATTCTG   480
GGGCCACGGT TGAGTTCAAC CCACTGACTT CAGGTGAAGG ACTGTGGACC AGCTTGAGAA   540
GAGGCCTCAC CAGAGTGGGT GTGGGGCATG GGGGCTCGAG CAGTACCCAG AGTAGGTGTG   600
GGTAGCCCGG CCAGGGGTTA ACGTGGGGCG TGGATTCAAC ACAGCTTGGA AGCCCAGAGC   660
TCGGAGGCCC GGGTGCTTGG GCCAATTGAG GAACAGGAGT CAGTCCATCC CGAGGGGGTT   720
GTCTCACTAC AATCTTCACA CGCCTTTATT ATTCACCATG GTTGGTGGCA CCTGGTTAGC   780
AGCAAGCGGA AGGCTGAGGC CAGTAGGGGC AGGGGTGTTA CTGGGGGTCG AAGAAGCCAG   840
CACAGAGACA GGGGTAGGGC CAGGGGTCGG GGCCACGGCC TGGATGAGGC CCACATGGGC   900
AGGCTGGCTG ATGAGATGGT GCTGCCCCCC TGCTGACACG AGGTGCACCA CATTCCTTTG   960
```

```
CAGCGGGCGG GCTGCCCCAC AGCAAGCTGG CGCACCTGGG CACCATCCAA AATACAGCTT  1020
GTTTCCCTGG ATTTGGAAGG TGAGAGGTTT GCTTCCCCCT CCATTAACCA CTGACGTTGT  1080
GCCAGTGAGA CTAACTCTCC GCGCCAATCT GTCCGCGGCT GACCTCCTTC GCGGGCGTGG  1140
CCTACCTCTT CCTCATGTTC CACACTGTCC CCGCACAGCC CGACTTTCAC TTGAGGGCTG  1200
GTTCCTGCGG CAGGGCTTGC TGGACACAAA CCTCACTGCG TCGGTGGCCA CACTGCTGGC  1260
CATCGCCGTG GAGCGGCACC GCAGTGTGAT GGCCGTGCAG CTGCACAGCC GCCTGCCCCG  1320
TGGCCGCGTG GTCATGCTCA TTGTGGGCGT GTGGGTGGCT GCCCTGGGCC TGGGGCTGCT  1380
GCCTGCCCAC TCCTGGCACT GCCTCTGTGC CCTGGACCGC TCCTCACGCA TGGCACCCCT  1440
GCTCAGCCGC TCCTATTTGG CCGTCTGGGC TCTGTCGAGC CTGCTTGTCT TCCTGCTCAT  1500
GGTGGCTGTG TACACCCGCA TTTTCTTCTA CGTGCGGCGG CGAGTGCAGC GCATGGCAGA  1560
GCATGTCAGC TGCCACCCCC GCTACCGAGA GACCACGCTC AGCCTGGTCA AGACTGTTGT  1620
CATCATCCTG GGGGCGTTCG TGGTCTGCTG GACACCAGGC CAGGTGGTAC TGCTCCTGGA  1680
TGGTTTAGGC TGTGAGTCCT GCAATGTCCT GGCGTTAGAA AAGTACTTCC TACTGTTGGC  1740
CGAGCCAACC TCACTGGTCA ATGCTGCTGT GTACTCTTGC CGAGATGCTG AGATGCGCCG  1800
CACCTTCCGC CGCCTTCTCC TGCTGCGCGT GCCTCCGCCA GTCCACCCGC GAGTCTGTCC  1860
ACTATACATC CTCTGCCCAG GGAGGTGCCA GCACTCGCAT CATGCTTCCC GAGAACGGCC  1920
ACCCACTGAT GGACTCCACC CTTTAGCTAC CTTGAACTAC AGCGGTACGC GGCAAGCAAC  1980
AAATCCACAG CCCCTGATGA CTTGTGGGTG CTCCTGGCTC AACCCAACCT CGTGCCGAAT  2040
TCCTGCAGCC CGGGGGATCC ACTAGTTCTA GAGCGGCGCC ACCGCGGTGG AGCTCCAGCT  2100
TTTGTTCCCT TTAGTGAGGG TTAATTTCGA GCTTGGCGTA ATCATGGTCA TAGCTGTTTC  2160
CTGTGTGAAA TTGTTATCCG CTCAC                                        2185
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS

(A) LENGTH: 393 AMINO ACIDS

(B) TYPE: AMINO ACID

(C) STRANDEDNESS:

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Arg Pro Thr Trp Ala Gly Trp Leu Met Arg Trp Cys Cys Pro
  5                  10                  15
Pro Ala Asp Thr Arg Cys Thr Thr Phe Leu Cys Ser Gly Arg Ala
            20                  25                  30
Ala Pro Gln Gln Ala Gly Ala Pro Gly His His Pro Lys Tyr Ser
        35                  40                  45
Leu Phe Pro Trp Ile Trp Lys Val Arg Gly Leu Leu Pro Pro Pro
    50                  55                  60
Leu Thr Thr Asp Val Val Pro Val Arg Leu Thr Leu Arg Ala Asn
65                  70                  75
Leu Ser Ala Ala Asp Leu Leu Arg Gly Arg Gly Leu Pro Leu Pro
            80                  85                  90
His Val Pro His Cys Pro Arg Thr Ala Arg Leu Ser Leu Glu Gly
        95                  100                 105
Trp Phe Leu Arg Gln Gly Leu Leu Asp Thr Asn Leu Thr Ala Ser
    110                 115                 120
Val Ala Thr Leu Leu Ala Ile Ala Val Glu Arg His Arg Ser Val
125                 130                 135
Met Ala Val Gln Leu His Ser Arg Leu Pro Arg Gly Arg Val Val
            140                 145                 150
Met Leu Ile Val Gly Val Trp Val Ala Ala Leu Gly Leu Gly Leu
        155                 160                 165
Leu Pro Ala His Ser Trp His Cys Leu Cys Ala Leu Asp Arg Ser
    170                 175                 180
```

```
Ser Arg Met Ala Pro Leu Leu Ser Arg Ser Tyr Leu Ala Val Trp
                185                 190                 195
Ala Leu Ser Ser L u Leu Val Phe Leu Leu Met Val Ala Val Tyr
                200                 205                 210
Thr Arg Ile Phe Phe Tyr Val Arg Arg Arg Val Gln Arg Met Ala
                215                 220                 225
Glu His Val Ser Cys His Pro Arg Tyr Arg Glu Thr Thr Leu Ser
                230                 235                 240
Leu Val Lys Thr Val Val Ile Ile Leu Gly Ala Phe Val Val Cys
                245                 250                 255
Trp Thr Pro Gly Gln Val Val Leu Leu Leu Asp Gly Leu Gly Cys
                260                 265                 270
Glu Ser Cys Asn Val Leu Ala Leu Glu Lys Tyr Phe Leu Leu Leu
                275                 280                 285
Ala Glu Pro Thr Ser Leu Val Asn Ala Ala Val Tyr Ser Cys Arg
                290                 295                 300
Asp Ala Glu Met Arg Arg Thr Phe Arg Arg Leu Leu Leu Leu Arg
                305                 310                 315
Val Pro Pro Pro Val His Pro Arg Val Cys Pro Leu Tyr Ile Leu
                320                 325                 330
Cys Pro Gly Arg Cys Gln His Ser His His Ala Ser Arg Glu Arg
                335                 340                 345
Pro Pro Thr Asp Gly Leu His Pro Leu Ala Thr Leu Asn Tyr Ser
                350                 355                 360
Gly Thr Arg Gln Ala Thr Asn Pro Gln Pro Leu Met Thr Cys Gly
                365                 370                 375
Cys Ser Trp Leu Asn Pro Thr Ser Cys Arg Ile Pro Ala Ala Arg
                380                 385                 390
Gly Ile His
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 1474 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CGGCACGAGC ATAAGAAGAC AGAGAGAACT GAGTATCCTC CCAAAGGTGA CACTGGAAGC   60
AATGAACACC ACAGTAATGC AAGGCTTGAA CAGATCTAAG CGGTGCCCCA AAGACACTCG  120
GATAGTACAG CTGGTATTCC CAGCCCTCTA CACAGTGGTT TTCTTGACCG CAATCCTGCT  180
GAATACTTTG GCTCTGTGGG TGTTTGTTCA CATCCCCAGC TGGTCCACCT TCATCATCTA  240
CCTCAAAAAC ACTTTGGTGG CCGACTTGAT AATGACAGTG ATGCTTCCTT TCAAAATCCT  300
CTCTGACTCA CACCTGGCAC CCTGGCAGCT CAGAGCTTTT GTGTGTCGTT TTTCTTCGGT  360
GATATTTTAT GAGACCATGT ATGTGGGCAT AGTGCTGTTA GGGCTCATAG CCTTTGACAG  420
ATTCCTCAAG ATCATCAGAC CTTTGAGAAA TATTTTTCTA AAAAAACCTG TTTGGGGAAA  480
AACGGTCTCA ATCTTCATCT GGTTCTTTTG GTTCTTCATC TCCCTGCCAA ATATGATCTT  540
GAGCAACAAG GAAGCAACAC CATCGTCTGT GAAAAAGTGT GCTTCCTTAA AGGGGCCTCT  600
GGGGCTGAAA TGGCATCAAA TGGTAAATAA CATATGCCAG TTTATTTTCT GGACTGTTTT  660
TATCCTAATG CTTGTGTTTT ATGTGGTTAT TGCAAAAAAG TATATGATTC TTATAGAAAG  720
TCCAAAAGTA AGGACAGAAA AAACAACAAA AAGCTGGAAG GCAAAGTATT TGTTGTCGTG  780
```

```
GCTGTCTTCT TTGTGTGTTT TGCTCCATTT CATTTCGCCA GAGTTCCATA TACTCACAGT   840
CAAACCAACA ATAAGACTGA CTGTAGACTG CAAAATCAAC TGTTTATTGC TAAAGAAACA   900
ACTCTCTTTT TGGCAGCAAC TAACATTTGT ATGGATCCCT TAATATACAT ATTCTTATGT   960
AAAAAATTCA CAGAAAAGCT ACCATGTATG CAAGGGAGAA AGACCACAGC ATCAAGCCAA  1020
GAAAATCATA GCAGTCAGAC AGACAACATA ACCTTAGGCT GACAACTGTA CATAGGGGTA  1080
ACTTCTATTT ATTGATGAGA CTTCCGTAGA TAATGTGGAA ATCCAATTTA ACCAAGAAAA  1140
AAAGATTGGG GCAAATGCTC TCTTACATTT TATTATCCTG GTGTACAGAA AAGATTATAT  1200
AAAATTTAAA TCCACATAGA TCTATTCATA AGCTGAATGA ACCATTACTA AGAGAATGCA  1260
ACAGGATACA AATGGCCACT AGAGGTCATT ATTTGTTTCT TTCTTTCTTT TTTTTTTTTT  1320
AATTTCAAGA GCATTTCACT TTAACATTTT GGAAAAGACT AAGGAGAAAC GTATATCCCT  1380
ACAAACCTCC CCTCCAAACA CCTTCTTACA TTCTTTTCCA CAATTCACAT AACACTACTG  1440
CTTTTGTGCC CCTTAAATGT AGATTTGTTG GCTG                              1474
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS
- (A) LENGTH: 293 AMINO ACIDS
- (B) TYPE: AMINO ACID
- (C) STRANDEDNESS:
- (D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Met Asn Thr Thr Val Met Gln Gly Phe Asn Arg Ser Lys Arg Cys
  5                  10                  15
Pro Lys Asp Thr Arg Ile Val Gln Leu Val Phe Pro Ala Leu Tyr
            20                  25                  30
Thr Val Val Phe Leu Thr Gly Ile Leu Leu Asn Thr Leu Ala Leu
        35                  40                  45
Trp Val Phe Val His Ile Pro Ser Ser Ser Thr Phe Ile Ile Tyr
    50                  55                  60
Leu Lys Asn Thr Leu Val Ala Asp Leu Ile Met Thr Leu Met Leu
65                  70                  75
Pro Phe Lys Ile Leu Ser Asp Ser His Leu Ala Pro Trp Gln Leu
            80                  85                  90
Arg Ala Phe Val Cys Arg Phe Ser Ser Val Ile Phe Tyr Glu Thr
        95                  100                 105
Met Tyr Val Gly Ile Val Leu Leu Gly Leu Ile Ala Phe Asp Arg
    110                 115                 120
Phe Leu Lys Ile Ile Arg Pro Leu Arg Asn Ile Phe Leu Lys Lys
125                 130                 135
Pro Val Trp Gly Lys Thr Val Ser Ile Phe Ile Trp Phe Phe Trp
            140                 145                 150
Phe Phe Ile Ser Leu Pro Asn Met Ile Leu Ser Asn Lys Glu Ala
        155                 160                 165
Thr Pro Ser Ser Val Lys Lys Cys Ala Ser Leu Lys Gly Pro Leu
    170                 175                 180
Gly Leu Lys Trp His Gln Met Val Asn Asn Ile Cys Gln Phe Ile
185                 190                 195
Phe Trp Thr Val Phe Ile Leu Met Leu Val Phe Tyr Val Val Ile
            200                 205                 210
Ala Lys Lys Tyr Met Ile Leu Ile Glu Ser Pro Lys Val Arg Thr
        215                 220                 225
Glu Lys Thr Thr Lys Ser Trp Lys Ala Lys Tyr Leu Leu Ser Trp
```

```
                    230                 235                 240
Leu Ser Ser Leu Cys Val Leu Leu His Phe Ile Ser Pro Glu Phe
                    245                 250                 255
His Ile Leu Thr Val Lys Pro Thr Ile Arg Leu Thr Val Asp Cys
                    260                 265                 270
Lys Ile Asn Cys Leu Leu Leu Lys Lys Gln Leu Ser Phe Trp Gln
                    275                 280                 285
Gln Leu Thr Phe Val Trp Ile Pro
                    290
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 1301 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
TTTTGGGTAT TTCTGAGAAA AAGGAAATAT TTATAAAACC ATCCAAAGAT CCAGATAATT    60
TGCAAATAAA TTGGAGGTTA TAGAGGTTAT AATCTGAATC CCAAAGGAGA CTGCAGCTGA   120
TGAAAGTGCT TCCAAACTGA AAATTGGACG TGCCTTTACG ATGGTAAGCG TTAACAGCTC   180
CCACTGCTTC TATAATGACT CCTTTAAGTA CACTTTGTAT GGGTGCATGT TCAGCATGGT   240
GTTTGTGCTT GGGTTAATAT CCAATTGTGT TGCCATATAC ATTTTCATCT GCGTCCTCAA   300
AGTCCGAAAT GAAACTACAA CTTACATGAT TAACTTGGCA ATGTCAGACT TGCTTTTTGT   360
TTTTACTTTA CCCTTCAGGA TTTTTTACTT CACAACACGG AATTGGCCAT TTGGAGATTT   420
ACTTTGTAAG ATTTCTGTGA TGCTGTTTTA TACCAACATG TACGGAAGCA TTCTGTTCTT   480
AACCTGTATT AGTGTAGATC GATTTCTGGC AATTGTCTAC CCATTTAAGT CAAAGACTCT   540
AAGAACCAAA AGAAATGCAA AGATTGTTTG ACATGGCGTG TGGTTAACTG TGATCGGAGG   600
AAGTGCACCC GCCGTTTTTG TTCAGTCTAC CCACTCTCAG GGTAACAATG CCTCAGAAGC   660
CTGCTTTGAA AATTTTCCAG AAGCCACATG GAAAACATAT CTCTCAAGGA TTGTAATTTT   720
CATCGAAATA GTGGGATTTT TTATTCCTCT AATTTTAAAT GTAACTTGTT CTAGTATGGT   780
GCTAAAAACT TTAACCAAAC CTGTTACATT AAGTAGAAGC AAAATAAACA AAACTAAGGT   840
TTTAAAAATG ATTTTTGTAC ATTTGATCAT ATTCTGTTTC TGTTTTGTTC CTTACAATAT   900
CAATCTTATT TTATATTCTC TTGTGAGAAC ACAAACATTT GTTAATTGCT CAGTAGTGGC   960
AGCAGTAAGG ACAATGTACC CAATCACTCT CTGTATTGCT GTTTCCAACT GTTGTTTTGA  1020
CCCTATAGTT TACTACTTTA CATCGGACAC AATTCAGAAT TCAATAAAAA TGAAAAACTG  1080
GTCTGTCAGG AGAAGTGACT TCAGATTCTC TGAAGTTCAT GGTGCAGAGA ATTTTATTCA  1140
GCATAACCTA CAGACCTTAA AAAGTAAGAT ATTTGACAAT GAATCTGCTG CCTGAAATAA  1200
AACCATTAGG ACTCACTGGG ACAGAACTTT CAAGTTCCTT CAACTGTGAA AAGTGTCTTT  1260
TTGGACAAAC TATTTTTCCA CCTCCAAAAG AAATTAACAC A                       1301
```

(2) INFORMATION FOR SEQ ID NO:8:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 344 AMINO ACIDS
(B) TYPE: AMINO ACID
(C) STRANDEDNESS:
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: PROTEIN

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Val Ser Val Asn Ser Ser His Cys Phe Tyr Asn Asp Ser Phe

```
                    5                  10                  15
Lys Tyr Thr L u Tyr Gly Cys Met Ph  S r M t Val Phe Val Leu
                20                  25                  30
Gly Leu Ile S r Asn Cys Val Ala Il  Tyr Ile Phe Il  Cys Val
            35                  40                  45
Leu Lys Val Arg Asn Glu Thr Thr Thr Tyr Met Ile Asn Leu Ala
        50                  55                  60
Met Ser Asp Leu Leu Phe Val Phe Thr Leu Pro Phe Arg Ile Phe
    65                  70                  75
Tyr Phe Thr Thr Arg Asn Trp Pro Phe Gly Asp Leu Leu Cys Lys
80                  85                  90
Ile Ser Val Met Leu Phe Tyr Thr Asn Met Tyr Gly Ser Ile Leu
            95                  100                 105
Phe Leu Thr Cys Ile Ser Val Asp Arg Phe Leu Ala Ile Val Tyr
        110                 115                 120
Pro Phe Lys Ser Lys Thr Leu Arg Thr Lys Arg Asn Ala Lys Ile
    125                 130                 135
Val Cys Thr Gly Val Trp Leu Thr Val Ile Gly Gly Ser Ala Pro
140                 145                 150
Ala Val Phe Val Gln Ser Thr His Ser Gln Gly Asn Asn Ala Ser
            155                 160                 165
Glu Ala Cys Phe Glu Asn Phe Pro Glu Ala Thr Trp Lys Thr Tyr
        170                 175                 180
Leu Ser Arg Ile Val Ile Phe Ile Glu Ile Val Gly Phe Phe Ile
    185                 190                 195
Pro Leu Ile Leu Asn Val Thr Cys Ser Ser Met Val Leu Lys Thr
200                 205                 210
Leu Thr Lys Pro Val Thr Leu Ser Arg Ser Lys Ile Asn Lys Thr
            215                 220                 225
Lys Val Leu Lys Met Ile Phe Val His Leu Ile Ile Phe Cys Phe
        230                 235                 240
Cys Phe Val Pro Tyr Asn Ile Asn Leu Ile Leu Tyr Ser Leu Val
    245                 250                 255
Arg Thr Gln Thr Phe Val Asn Cys Ser Val Val Ala Ala Val Arg
260                 265                 270
Thr Met Tyr Pro Ile Thr Leu Cys Ile Ala Val Ser Asn Cys Cys
            275                 280                 285
Phe Asp Pro Ile Val Tyr Tyr Phe Thr Ser Asp Thr Ile Gln Asn
        290                 295                 300
Ser Ile Lys Met Lys Asn Trp Ser Val Arg Arg Ser Asp Phe Arg
    305                 310                 315
Phe Ser Glu Val His Gly Ala Glu Asn Phe Ile Gln His Asn Leu
320                 325                 330
Gln Thr Leu Lys Ser Lys Ile Phe Asp Asn Glu Ser Ala Ala
            335                 340
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
GACTAAAGCT TAATGAGTAG TGAAATGGTG 30

(2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 31 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
GAACTTCTAG ACCCTCAGGG TTGTAAATCA G 31

(2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
GACTAAAGCT TAATGAGGCC CACATGGGCA 30

(2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 32 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GAACTTCTAG ACGAACTAGT GGATCCCCCC GG 32

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GACTAAAGCT TAATGGCGTC TTTCTCTGCT 30

(2) INFORMATION FOR SEQ ID NO:14:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:
GAACTTCTAG ACTTCACACA GTTGTACTAT 30

(2) INFORMATION FOR SEQ ID NO:15:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:
GACTAAAGCT TAATGGTAAG CGTTAACAGC 30

(2) INFORMATION FOR SEQ ID NO:16:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 31 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:
GAACTTCTAG ACTTCAGGCA GCAGATTCAT T 31

(2) INFORMATION FOR SEQ ID NO:17:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 34 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:
GTCCAAGCTT GCCACCATGA GTAGTGAAAT GGTG 34

(2) INFORMATION FOR SEQ ID NO:18:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 58 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE

(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:
CTAGCTCGAG TCAAGCGTAG TCTGGGACGT CGTATGGGTA GCAGGGTTGT AAATCAGG 58

(2) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 34 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
GTCCAAGCTT GCCACCATGG TTGGTGGCAC CTGG 34

(2) INFORMATION FOR SEQ ID NO:20:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 58 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
CTAGCTCGAG TCAAGCGTAG TCTGGGACGT CGTATGGGTA GCAGTGGATC CCCCGTGC 58

(2) INFORMATION FOR SEQ ID NO:21:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 34 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:
GTCCAAGCTT GCCACCATGA ACACCACAGT AATG 34

(2) INFORMATION FOR SEQ ID NO:22:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 61 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CTAGCTCGAG TCAAGCGTAG TCTGGGACGT CGTATGGGTA GCAAGGGATC CATACAAATG 60
T 61

(2) INFORMATION FOR SEQ ID NO:23:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 34 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
GTCCAAGCTT GCCACCATGG TAAGCGTTAA CAGC 34

(2) INFORMATION FOR SEQ ID NO:24:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 61 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
CTAGCTCGAG TCAAGCGTAG TCTGGGACGT CGTATGGGTA GCAGGCAGCA GATTCATTGT 60
C 61

(2) INFORMATION FOR SEQ ID NO:25:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 30 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:
CGGGATCCCT CCATGAGTAG TGAAATGGTG 30

(2) INFORMATION FOR SEQ ID NO:26:
(i) SEQUENCE CHARACTERISTICS
(A) LENGTH: 29 BASE PAIRS
(B) TYPE: NUCLEIC ACID
(C) STRANDEDNESS: SINGLE
(D) TOPOLOGY: LINEAR

(ii) MOLECULE TYPE: Oligonucleotide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:
CGGGATCCCG CTCAGGGTTG TAAATCAGG 29

WHAT IS CLAIMED IS:

1. An isolated polynucleotide selected from th group consisting of:

(a) a polynucleotide encoding the polypeptide as set forth in SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6, and SEQ ID No. 8 or fragments, analogs or derivatives of said polypeptides;

(b) a polynucleotide capable of hybridizing to and which is at least 70% identical to the polynucleotide of (a); and

(c) a fragment of the polynucleotides of (a) or (b) wherein said fragment has at least 50 nucleotides.

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 1 wherein the polynucleotide is genomic DNA.

5. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) the polynucleotide of Claim 2 encoding a polypeptide having the amino acid sequence encoded by the DNA contained in ATCC Deposit No. 75981;

(b) a polynucleotide capable of hybridizing to and which is at least 70% identical to the polynucleotide of (a); and

(c) a fragment of the polynucleotides of (a) or (b) wherein said fragment has at least 50 nucleotides.

6. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) the polynucleotide of Claim 2 encoding a polypeptide having the amino acid sequence encoded by the DNA contained in ATCC Deposit No. 75983;

(b) a polynucl otide capable of hybridizing to and which is at least 70% identical to the polynucleotide of (a); and

(c) a fragment of the polynucleotides of (a) or (b) wherein said fragment has at least 50 nucleotides.

7. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) the polynucleotide of Claim 2 encoding a polypeptide having the amino acid sequence encoded by the DNA contained in ATCC Deposit No. 75967;

(b) a polynucleotide capable of hybridizing to and which is at least 70% identical to the polynucleotide of (a); and

(c) a fragment of the polynucleotides of (a) or (b) wherein said fragment has at least 50 nucleotides.

8. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) the polynucleotide of Claim 2 encoding a polypeptide having the amino acid sequence encoded by the DNA contained in ATCC Deposit No. 75979;

(b) a polynucleotide capable of hybridizing to and which is at least 70% identical to the polynucleotide of (a); and

(c) a fragment of the polynucleotides of (a) or (b) wherein said fragment has at least 50 nucleotides.

9. The polynucleotide of Claim 1 encoding the polypeptide having the amino acid sequence as shown in SEQ ID No. 2.

10. The polynucleotide of claim 9 having the coding sequence as shown in SEQ ID No. 1 from nucleotide 1 to nucleotide 1713.

11. The polynucleotide of Claim 1 encoding the polypeptide having the amino acid sequence as shown in SEQ ID No. 4.

12. The polynucleotide of claim 11 having the coding sequence as shown in SEQ ID No. 3 from nucl otide 1 to nucleotide 2185.

13. The polynucleotide of Claim 1 encoding the polypeptide having the amino acid sequence as shown in SEQ ID No. 6.

14. The polynucleotide of claim 13 having the coding sequence as shown in SEQ ID No. 5 from nucleotide 1 to nucleotide 1474.

15. The polynucleotide of Claim 1 encoding the polypeptide having the amino acid sequence as shown in SEQ ID No. 8.

16. The polynucleotide of claim 15 having the coding sequence as shown in SEQ ID No. 7 from nucleotide 1 to nucleotide 1301.

17. A vector containing the DNA of Claim 2.

18. A host cell genetically engineered with the vector of Claim 17.

19. A process for producing a polypeptide comprising: expressing from the host cell of Claim 18 the polypeptide encoded by said DNA.

20. A process for producing cells capable of expressing a polypeptide comprising genetically engineering cells with the vector of Claim 17.

21. An isolated DNA hybridizable to the DNA of Claim 2 and encoding a polypeptide having G-protein coupled receptor activity.

22. A polypeptide selected from the group consisting of: (i) a polypeptide having the deduced amino acid sequence of SEQ ID No. 2, SEQ ID No. 4, SEQ ID No. 6 and SEQ ID No. 8 and fragments, analogs and derivatives thereof, (ii) a polypeptide encoded by the cDNA of ATCC Deposit No. 75981, ATCC Deposit No. 75983, ATCC Deposit No. 75976 and ATCC Deposit No. 75979 and fragments, analogs and derivatives of said polypeptide.

23. An antibody against the polypeptide of claim 22.

24. A compound which activates the polypeptide of Claim 22.

25. A compound which inhibits activation of the polypeptide of claim 22.

26. A method for the treatment of a patient having need to activate a G-protein coupled receptor comprising: administering to the patient a therapeutically effective amount of the compound of Claim 24.

27. A method for the treatment of a patient having need to inhibit activation of a G-protein coupled receptor comprising: administering to the patient a therapeutically effective amount of the compound of Claim 25.

28. The polypeptide of Claim 22 wherein the polypeptide is a soluble fragment of the G-protein coupled receptor and is capable of binding a ligand for the receptor.

29. A process for identifying antagonists and agonists to the polypeptide of claim 22 comprising:

contacting a cell which expresses a G-protein coupled receptor with a known receptor ligand and a compound to be screened; and

determining if the compound inhibits or enhances activation of the receptor.

30. A process for determining whether a ligand not known to be capable of binding to the polypeptide of claim 22 can bind thereto comprising:

contacting a mammalian cell which expresses a G-protein coupled receptor with a potential ligand;

detecting the presence of the ligand which binds to the receptor; and

determining whether the ligand binds to the G-protein coupled receptor.

31. A method for diagnosing a disease or a susceptibility to a disease comprising:

detecting a mutation in the nucleic acid sequence encoding the polypeptide of claim 22 in a sample derived from a host.

32. A diagnostic process comprising:

analyzing for the presence of the polypeptide of claim 28 in a sample derived from a host.

FIG. 1A

```
        10                  30                  50
GGCACGAGGTCATTCAACATTTATTCAACCAAAAATACTAAGTCAGCTCTATACAAACTA
        70                  90                  110
ATGGAAGGATACAGCTATGCAAATATAGAACACTAAAGTGTTACATGACAGATGTATGAG
                                                    M  S
       130                 150                 170
TAGTGAAATGGTGAAAAATCAGACAATGGTCACAGAGTTCCTCCTACTGGGATTTCTCCT
 S  E  M  V  K  N  Q  T  M  V  T  E  F  L  L  L  G  F  L  L
       190                 210                 230
GGGCCCAAGGATTCAGATGCTCCTCTTTGGGCTCTTCTCCCTGTTCTATGTCTTCACCCT
 G  P  R  I  Q  M  L  L  F  G  L  F  S  L  F  Y  V  F  T  L
       250                 270                 290
GCTGGGGAATGGGACCATCCTGGGGCTCATCTCACTGGACTCCAGACTCCACACCCCCAT
 L  G  N  G  T  I  L  G  L  I  S  L  D  S  R  L  H  T  P  M
       310                 330                 350
GTACTTCTTCCTCTCACACCTGGCCGTCGTCAACATCGCCTATGCCTGCAACACAGTGCC
 Y  F  F  L  S  H  L  A  V  V  N  I  A  Y  A  C  N  T  V  P
       370                 390                 410
CCAGATGCTGGTGAACCTCCTGCATCCAGCCAAGCCCATCTCCTTTGCTGGTTGCATGAC
 Q  M  L  V  N  L  L  H  P  A  K  P  I  S  F  A  G  C  M  T
       430                 450                 470
ACTAGACTTTCTCTTTTTGAGTTTTGCACATACTGAATGCCTCCTGTTGGTGCTGATGTC
 L  D  F  L  F  L  S  F  A  H  T  E  C  L  L  L  V  L  M  S
       490                 510                 530
CTACGATCGGTACGTGGCCATCTGCCACCCTCTCCGATATTTCATCATCATGACCTGGAA
 Y  D  R  Y  V  A  I  C  H  P  L  R  Y  F  I  I  M  T  W  K
       550                 570                 590
AGTCTGCATCACTCTGGGCATCACTTCCTGGACATGTGGCTCCCTCCTGGCTATGGTCCA
```

MATCH WITH FIG. 1B

FIG. 1B

```
MATCH WITH FIG. 1A

V C I T L G I T S W T C G S L L A M V H
      610       630       650
TGTGAGCCTCATCCTAAGACTGCCCTTTGTGGGCCTCGTGAAATCAACCACTTCTTCTG
 V S L I L R L P F C G P R E I N H F F C
      670       690       710
TGAAATCCTGTCTGTCCTCAGGCTGGCCTGTGCTGATACCTGGCTCAACCAGGTGGTCAT
 E I L S V L R L A C A D T W L N Q V V I
      730       750       770
CTTTGaAGCCTGCATGTTCATCCTGGTGGGACCACTCTGCCTGGTGCTGGTCTCCTACTC
 F E A C M F I L V G P L C L V L V S Y S
      790       810       830
ACACATCCTGGGGGGCATCCTGAGGATCCAGTCTGGGGAGGGCCGCAGAAAGGCCTTCTC
 H I L G G I L R I Q S G E G R R K A F S
      850       870       890
CACCTGCTCCTCCCACCTCTGCGTAGTGGGACTCTTCTTTGGSAGCGCCATCGTCATGTA
 T C S S H L C V V G L F F G S A I V M Y
      910       930       950
CATGGCCCCTAAGTCCCGCCATCCTGAGGAGCAGCAGAAGGTCCtTTTTCtTATTTTACA
 M A P K S R H P E E Q Q K V L F L I L Q
      970       990      1010
GTTCcTTTCAACCCCGATGcTTAAACCCCCCTGATTTACAACCCTGAGGAATGTAGAGGGT
 F L S T P M L K P P D L Q P *
     1030      1050      1070
CAaGggtGCcCTCCGAGGAGACCActGTGCAARGRAAGTCAtTCCtAAGGGGTGTGACAT
     1090      1110      1130
          MATCH WITH FIG. 1C
```

FIG. IC

MATCH WITH FIG. IB

```
TTGAACTGCCAGCCCCAGTTGCCCCGTGGACTCCTGATGCCCAATTATTGCCTCAACCCA
      1150                1170                1190
GAAAAGTTTACTCCCCTTTAACTGTGCTTTACTGACAGAAGGGCAAGCCTTCTCCCGTTT
      1210                1230                1250
TTTGCAGATAAAATTTTAGATGTGTTGCAATCATTGGGTTTCTAGGAGATGTGGTTTTAT
      1270                1290                1310
CAGACAATTTTTTCTTTTATTTCACAATTACTTTAATATCTGTAAAATAAAGAATTATTT
      1330                1350                1370
TAAATCATTTTCCCAGTCCCAAAAGTTAAATACAGGCCACTTACTTCTTTAACCAAATGA
      1390                1410                1430
TATAGTTTGGCTCTGTGTCCCCACCCAAATCTCATGTCAAATTGTAATCCCCGCATGTCA
      1450                1470                1490
GCGGAGGGACCTGGTGGGAGGTGATTGGATCATGGGGAGGGATTTCCCCCTTGCTGTTCT
      1510                1530                1550
GTTGATAGTGAACGAGTTCTCACGAAATCTGATGGTTTAAAAGTGCAGCACTTCTCCCTT
      1570                1590                1610
TGCTCTCTCTCTCCTGCTGTGCCATGGTAAGACGTGCCTTGCTTCCCCTGGTGCTTCCGC
      1630                1650                1670
CATGATTGTACCTTTCCTGAGGCCTCTCCAGCCATGTGGAACTGTGAGCCAATTAAACTT
      1690                1710
CTTTTCTTTAGAAAAAAAAAAAAAAAAAAAA
```

FIG. 2A

```
          10                            30                            50
TCACTATAGGGCGAATTGGGTACGGGCCCCCCCTCGAGGTCGACGGTATCGATAAGCTTG
          70                            90                           110
ATATCGAATTCGGCACGAGCCGGGCTCGGAGAGGTGACGGAACCGGGGCTGGTAGCATAG
         130                           150                           170
TTTGATTTGATGATGGAGCCAACACAGGGGTTGGAGCTGGTACCGGTGAAGCTGAGGCTA
         190                           210                           230
AAAAGGTTCCTGGAGTAGACGATGGAGCCATAACTGGAACCGGAGTCTGTGAATGAAGCC
         250                           270                           290
AGGACAGGAGCAGCACCTGGCGATGGTGCCAGGACCGGAAGAGGAGCCAGAGGAGGAGCT
         310                           330                           350
GGAGAAGGAGCCAGAATTGCTGTCTGTGGAGCCGCCATAGGAGCCAGAGGGGTGGCTAGA
         370                           390                           410
GCCTGAGAATGCAGAAGATGCTGGAGCCAGAAGGGAAGCCTGAGCTGGAGCTGGATTTGG
         430                           450                           470
TGCTGACGGAAAAGGACTGGCCAGAGCCGAAGCTGGCACCAGGGACAGGTGAGCATTCTG
         490                           510                           530
GGGCCACGGTTGAGTTCAACCCACTGACTTCAGGTGAAGGACTGTGGACCAGCTTGAGAA
         550                           570                           590
GAGGCCTCACCAGAGTGGGTGTGGGCATGGGGGCTCGAGCAGTACCCAGAGTAGGTGTG
         610                           630                           650
GGTAGCCCGGCCAGGGGTTAACGTGGGGCGTGGATTCAACACAGCTTGGAAGCCCAGAGC
         670                           690                           710
TCGGAGGCCCGGGTGCTTGGCCAATTGAGGAACAGGAGTCAGTCCATCCCGAGGGGGTT
         730                           750                           770
GTCTCACTACAATCTTCACACGCCTTTATTATTCACCATGGTTGGTGGCACCTGGTTAGC
         790                           810                           830
AGCAAGCGGAAGGCTGAGGCCAGTAGGGGCAGGGGTGTTACTGGGGGTCGAAGAAGCCAG
```

MATCH WITH FIG. 2B

FIG. 2B

MATCH WITH FIG. 2A

```
850                 870                 890
CACAGAGACAGGGGTAGGGCCAGGGGTCGGGGCCACGGCCTGGATGAGGCCCACATGGGC
                                          M  R  P  T  W  A
910                 930                 950
AGGCTGGCTGATGAGATGGTGCTGCCCCCCTGCTGACACGAGGTGCACCACATTCCTTTG
 G  W  L  M  R  W  C  C  P  P  A  D  T  R  C  T  T  F  L  C
970                 990                 1010
CAGCGGGCGGGCTGCCCCACAGCAAGCTGGCGCACCTGGGCACCATCCAAAATACAGCTT
 S  G  R  A  A  P  Q  Q  A  G  A  P  G  H  H  P  K  Y  S  L
1030                1050                1070
GTTTCCCTGGATTTGGAAGGTGAGAGGTTTGCTTCCCCCTCCATTAACCACTGACGTTGT
 F  P  W  I  W  K  V  R  G  L  L  P  P  P  L  T  T  D  V  V
1090                1110                1130
GCCAGTGAGACTAACTCTCCGCGCCAATCTGTCCGCGGCTGACCTCCTTCGCGGGCGTGG
 P  V  R  L  T  L  R  A  N  L  S  A  A  D  L  L  R  G  R  G
1150                1170                1190
CCTACCTCTTCCTCATGTTCCACACTGTCCCCGCACAGCCCGACTTTCACTTGAGGGCTG
 L  P  L  P  H  V  P  H  C  P  R  T  A  R  L  S  L  E  G  W
1210                1230                1250
GTTCCTGCGGCAGGGCTTGCTGGACACAAACCTCACTGCGTCGGTGGCCACACTGCTGGC
 F  L  R  Q  G  L  L  D  T  N  L  T  A  S  V  A  T  L  L  A
1270                1290                1310
CATCGCCGTGGAGCGGCACCGCAGTGTGATGGCCGTGCAGCTGCACAGCCGCCTGCCCCG
 I  A  V  E  R  H  R  S  V  M  A  V  Q  L  H  S  R  L  P  R
1330                1350                1370
TGGCCGCGTGGTCATGCTCATTGTGGGCGTGTGGGTGGCTGCCCTGGGCCTGGGGCTGCT
 G  R  V  V  M  L  I  V  G  V  W  V  A  A  L  G  L  G  L  L
1390                1410                1430
```

MATCH WITH FIG. 2C

FIG. 2C

MATCH WITH FIG. 2B

```
GCCTGCCCACTCCTGGCACTGCCTCTGTGCCCTGGACCGCTCCTCACGCATGGCACCCCT
 P  A  H  S  W  H  C  L  C  A  L  D  R  S  S  R  M  A  P  L
      1450                  1470                  1490
GCTCAGCCGCTCCTATTTGGCCGTCTGGGCTCTGTCGAGCCTGCTTGTCTTCCTGCTCAT
 L  S  R  S  Y  L  A  V  W  A  L  S  S  L  L  V  F  L  L  M
      1510                  1530                  1550
GGTGGCTGTGTACACCCGCATTTTCTTCTACGTGCGGCGGCGAGTGCAGCGCATGGCAGA
 V  A  V  Y  T  R  I  F  F  Y  V  R  R  R  V  Q  R  M  A  E
      1570                  1590                  1610
GCATGTCAGCTGCCACCCCCGCTACCGAGAGACCACGCTCAGCCTGGTCAAGACTGTTGT
 H  V  S  C  H  P  R  Y  R  E  T  T  L  S  L  V  K  T  V  V
      1630                  1650                  1670
CATCATCCTGGGGGCGTTCGTGGTCTGCTGGACACCAGGCCAGGTGGTACTGCTCCTGGA
 I  I  L  G  A  F  V  V  C  W  T  P  G  Q  V  V  L  L  L  D
      1690                  1710                  1730
TGGTTTAGGCTGTGAGTCCTGCAATGTCCTGGCGTTAGAAAAGTACTTCCTACTGTTGGC
 G  L  G  C  E  S  C  N  V  L  A  L  E  K  Y  F  L  L  L  A
      1750                  1770                  1790
CGAGCCAACCTCACTGGTCAATGCTGCTGTGTACTCTTGCCGAGATGCTGAGATGCGCCG
 E  P  T  S  L  V  N  A  A  V  Y  S  C  R  D  A  E  M  R  R
      1810                  1830                  1850
CACCTTCCGCCGCCTTCTCCTGCTGCGCGTGCCTCCGCCAGTCCACCCGCGAGTCTGTCC
 T  F  R  R  L  L  L  L  R  V  P  P  P  V  H  P  R  V  C  P
```

Match with FIG. 2D

SUBSTITUTE SHEET (RULE 26)

FIG. 2D

MATCH WITH FIG. 2C

```
1870                1890                1910
ACTATACATCCTCTGCCCAGGGAGGTGCCAGCACTCGCATCATGCTTCCCGAGAACGGCC
  L  Y  I  L  C  P  G  R  C  Q  H  S  H  H  A  S  R  E  R  P
      1930                1950                1970
ACCCACTGATGGACTCCACCCTTTAGCTACCTTGAACTACAGCGGTACGCGGCAAGCAAC
  P  T  D  G  L  H  P  L  A  T  L  N  Y  S  G  T  R  Q  A  T
      1990                2010                2030
AAATCCACACAGCCCCCTGATGACTTGTGGGTGCTCCTGGCTCAACCCAACCTCGTGCCGAAT
  N  P  Q  P  L  M  T  C  G  C  S  W  L  N  P  T  S  C  R  I
      2050                2070                2090
TCCTGCAGCCCCGGGGGATCCACTAGTTCTAGAGCGGCGCCACCGCGGTGGAGCTCCAGCT
  P  A  A  R  G  I  H  *
      2110                2130                2150
TTTGTTCCCTTTAGTGAGGGTTAATTTCGAGCTTGGCGTAATCATGGTCATAGCTGTTTC
      2170
CTGTGTGAAATTGTTATCCGCTCAC
```

FIG. 3A

```
          10                  30                  50
CGGCACGAGCATAAGAAGACAGAGAGAACTGAGTATCCTCCCAAAGGTGACACTGGAAGC
          70                  90                 110
AATGAACACCACAGTAATGCAAGGCTTCAACAGATCTAAGCGGTGCCCCAAAGACACTCG
 M  N  T  T  V  M  Q  G  F  N  R  S  K  R  C  P  K  D  T  R
         130                 150                 170
GATAGTACAGCTGGTATTCCCAGCCCTCTACACAGTGGTTTTCTTGACCGGAATCCTGCT
 I  V  Q  L  V  F  P  A  L  Y  T  V  V  F  L  T  G  I  L  L
         190                 210                 230
GAATACTTTGGCTCTGTGGGTGTTTGTTCACATCCCCAGCTCCTCCACCTTCATCATCTA
 N  T  L  A  L  W  V  F  V  H  I  P  S  S  S  T  F  I  I  Y
         250                 270                 290
CCTCAAAAACACTTTGGTGGCCGACTTGATAATGACACTCATGCTTCCTTTCAAAATCCT
 L  K  N  T  L  V  A  D  L  I  M  T  L  M  L  P  F  K  I  L
         310                 330                 350
CTCTGACTCACACCTGGCACCCTGGCAGCTCAGAGCTTTTGTGTGTCGTTTTTCTTCGGT
 S  D  S  H  L  A  P  W  Q  L  R  A  F  V  C  R  F  S  S  V
         370                 390                 410
GATATTTTATGAGACCATGTATGTGGGCATCGTGCTGTTAGGGCTCATAGCCTTTGACAG
 I  F  Y  E  T  M  Y  V  G  I  V  L  L  G  L  I  A  F  D  R
         430                 450                 470
ATTCCTCAAGATCATCAGACCTTTGAGAAATATTTTTCTAAAAAAACCTGTTTGGGGAAA
 F  L  K  I  I  R  P  L  R  N  I  F  L  K  K  P  V  W  G  K
         490                 510                 530
AACGGTCTCAATCTTCATCTGGTTCTTTTGGTTCTTCATCTCCCTGCCAAATATGATCTT
 T  V  S  I  F  I  W  F  F  W  F  F  I  S  L  P  N  M  I  L
         550                 570                 590
```

MATCH WITH FIG. 3B

MATCH WITH FIG. 3A

FIG. 3B

```
GAGCAACAAGGAAGCAACACCATCGTCTGTGAAAAAGTGTGCTTCCTTAAAGGGGCCTCT
 S  N  K  E  A  T  P  S  S  V  K  K  C  A  S  L  K  G  P  L
       610                  630                  650
GGGGCTGAAATGGCATCAAATGGTAAATAACATATGCCAGTTTATTTTCTGGACTGTTTT
 G  L  K  W  H  Q  M  V  N  N  I  C  Q  F  I  F  W  T  V  F
       670                  690                  710
TATCCTAATGCTTGTGTTTTATGTGGTTATTGCAAAAAAGTATATGATTCTTATAGAAAG
 I  L  M  L  V  F  Y  V  V  I  A  K  K  Y  M  I  L  I  E  S
       730                  750                  770
TCCAAAAGTAAGGACAGAAAAAACAACAAAAAGCTGGAAGGCAAAGTATTTGTTGTCGTG
 P  K  V  R  T  E  K  T  T  K  S  W  K  A  K  Y  L  L  S  W
       790                  810                  830
GCTGTCTTCTTTGTGTGTTTTGCTCCATTTCATTTCGCCAGAGTTCCATATACTCACAGT
 L  S  S  L  C  V  L  L  H  F  I  S  P  E  F  H  I  L  T  V
       850                  870                  890
CAAACCAACAATAAGACTGACTGTAGACTGCAAAATCAACTGTTTATTGCTAAAGAAACA
 K  P  T  I  R  L  T  V  D  C  K  I  N  C  L  L  L  K  K  Q
       910                  930                  950
ACTCTCTTTTTGGCAGCAACTAACATTTGTATGGATCCCTTAATATACATATTCTTATGT
 L  S  F  W  Q  Q  L  T  F  V  W  I  P  *
       970                  990                 1010
AAAAAATTCACAGAAAAGCTACCATGTATGCAAGGGAGAAAGACCACAGCATCAAGCCAA
     1030                 1050                 1070
GAAAATCATAGCAGTCAGACAGACAACATAACCTTAGGCTGACAACTGTACATAGGGGTA
```

MATCH WITH FIG. 3C

FIG. 3C

Match with FIG. 3B

```
1090                1110                1130
ACTTCTATTTATTGATGAGACTTCCGTAGATAATGTGGAAATCCAATTTAACCAAGAAAA
1150                1170                1190
AAAGATTGGGGGCAAATGCTCTCTTACATTTTATTATCCTGGTGTACAGAAAAGATTATAT
1210                1230                1250
AAAATTTAAATCCACATAGATCTATTCATAAGCTGAATGAACCATTACTAAGAGAATGCA
1270                1290                1310
ACAGGATACAAATGGCCACTAGAGGTCATTATTTCTTTCTTTCTTTTTTTTTTTTTTTT
1330                1350                1370
AATTTCAAGAGCATTTCACTTTAACATTTTGGAAAAGACTAAGGAGAAACGTATATCCCT
1390                1410                1430
ACAAACCTCCCCCTCCAAACACCTTCTTACATTCTTTTCCACAATTCACATAACACTACTG
1450                1470
CTTTTGTGCCCCTTAAATGTAGATTTGTTGGCTG
```

FIG. 4A

```
         10                  30                  50
TTTTGGGTATTTCTGAGAAAAAGGAAATATTTATAAAACCATCCAAAGATCCAGATAATT
         70                  90                 110
TGCAAATAAATTGGAGGTTATAGAGGTTATAATCTGAATCCCAAAGGAGACTGCAGCTGA
        130                 150                 170
TGAAAGTGCTTCCAAACTGAAAATTGGACGTGCCTTTACGATGGTAAGCGTTAACAGCTC
                                          M  V  S  V  N  S  S
        190                 210                 230
CCACTGCTTCTATAATGACTCCTTTAAGTACACTTTGTATGGGTGCATGTTCAGCATGGT
 H  C  F  Y  N  D  S  F  K  Y  T  L  Y  G  C  M  F  S  M  V
        250                 270                 290
GTTTGTGCTTGGGTTAATATCCAATTGTGTTGCCATATACATTTTCATCTGCGTCCTCAA
 F  V  L  G  L  I  S  N  C  V  A  I  Y  I  F  I  C  V  L  K
        310                 330                 350
AGTCCGAAATGAAACTACAACTTACATGATTAACTTGGCAATGTCAGACTTGCTTTTTGT
 V  R  N  E  T  T  T  Y  M  I  N  L  A  M  S  D  L  L  F  V
        370                 390                 410
TTTTACTTTACCCTTCAGGATTTTTTACTTCACAACACGGAATTGGCCATTTGGAGATTT
 F  T  L  P  F  R  I  F  Y  F  T  T  R  N  W  P  F  G  D  L
        430                 450                 470
ACTTTGTAAGATTTCTGTGATGCTGTTTTATACCAACATGTACGGAAGCATTCTGTTCTT
 L  C  K  I  S  V  M  L  F  Y  T  N  M  Y  G  S  I  L  F  L
        490                 510                 530
AACCTGTATTAGTGTAGATCGATTTCTGGCAATTGTCTACCCATTTAAGTCAAAGACTCT
 T  C  I  S  V  D  R  F  L  A  I  V  Y  P  F  K  S  K  T  L
```

MATCH WITH FIG. 4B

MATCH WITH FIG. 4A

FIG. 4B

```
         550                           570                           590
AAGAACCAAAAGAAATGCAAAGATTGTTTGCACTGGCGTGTGGTTAACTGTGATCGGAGG
 R  T  K  R  N  A  K  I  V  C  T  G  V  W  L  T  V  I  G  G
         610                           630                           650
AAGTGCACCCGCCGTTTTTGTTCAGTCTACCCACTCTCAGGGTAACAATGCCTCAGAAGC
 S  A  P  A  V  F  V  Q  S  T  H  S  Q  G  N  N  A  S  E  A
         670                           690                           710
CTGCTTTGAAAATTTTCCAGAAGCCACATGGAAAACATATCTCTCAAGGATTGTAATTTT
 C  F  E  N  F  P  E  A  T  W  K  T  Y  L  S  R  I  V  I  F
         730                           750                           770
CATCGAAATAGTGGGATTTTTTATTCCTCTAATTTTAAATGTAACTTGTTCTAGTATGGT
 I  E  I  V  G  F  F  I  P  L  I  L  N  V  T  C  S  S  M  V
         790                           810                           830
GCTAAAAACTTTAACCAAACCTGTTACATTAAGTAGAAGCAAAATAAACAAAACTAAGGT
 L  K  T  L  T  K  P  V  T  L  S  R  S  K  I  N  K  T  K  V
         850                           870                           890
TTTAAAAATGATTTTTGTACATTTGATCATATTCTGTTTCTGTTTTGTTCCTTACAATAT
 L  K  M  I  F  V  H  L  I  I  F  C  F  C  F  V  P  Y  N  I
         910                           930                           950
CAATCTTATTTTATATTCTCTTGTGAGAACACAAACATTTGTTAATTGCTCAGTAGTGGC
 N  L  I  L  Y  S  L  V  R  T  Q  T  F  V  N  C  S  V  V  A
```

MATCH WITH FIG. 4C

FIG. 4C

Match with FIG. 4B

```
     970                990                1010
AGCAGTAAGGACAATGTACCCAATCACTCTCTGTATTGCTGTTTCCAACTGTTGTTTTGA
 A  V  R  T  M  Y  P  I  T  L  C  I  A  V  S  N  C  C  F  D
    1030               1050               1070
CCCTATAGTTTACTACTTTACATCGGACACAATTCAGAATTCAATAAAATGAAAAACTG
 P  I  V  Y  Y  F  T  S  D  T  I  Q  N  S  I  K  M  K  N  W
    1090               1110               1130
GTCTGTCAGGAGAAGTGACTTCAGATTCTCTGAAGTTCATGGTGCAGAGAATTTTATTCA
 S  V  R  R  S  D  F  R  F  S  E  V  H  G  A  E  N  F  I  Q
    1150               1170               1190
GCATAACCTACAGACCTTAAAAAGTAAGATATTTGACAATGAATCTGCTGCCTGAAATAA
 H  N  L  Q  T  L  K  S  K  I  F  D  N  E  S  A  A  *
    1210               1230               1250
AACCATTAGGACTCACTGGGACAGAACTTTCAAGTTCCTTCAACTGTGAAAGTGTCTTT
    1270               1290
TTGGACAAACTATTTTTCCACCTCCAAAAGAAATTAACACA
```

FIG. 5

```
 64 FFLSHLAVVNIAYACNTVPQMLVNLLHPAKPISFAGCMTLDFLFLSFAHT 113
    |||||||:|:|||||||||||||||.|.||||:||||| .||||.||  |
  1 FFLSHLAIVDIAYACNTVPQMLVNLLDPVKPISYAGCMTQTFLFLTFAIT 50

114 ECLLLVLMSYDRYVAICHPLRYFIMTWKVCITLGITSWTCGSLLAMVHV 163
    ||||||:|||||||||||||||||  ||.|:|| |:::|||..| ||.::|:
 51 ECLLLVVMSYDRYVAICHPLRYSAIMSWRVCSTMAVTSWIIGVLLSLIHL 100

164 SLILRLPFCGPREINHFFCEILSVLRLACADTWLNQVVIFEACMFILVGP 213
    |:|.||||....:|||||||| .:|:|||||| ||:.:::.:.: :||||
101 VLLLPLPFCVSQKVNHFFCEITAILKLACADTHLNETMVLAGAVSVLVGP 150

214 LCLVLVSYSHILGGILRIQSGEGRRKAFSTCSSHLCVVGLFFGSAIVMYM 263
    :: ::|||. |||:||:|||:||.|||||||||||||||:|.||||||:
151 FSSIVVSYACILGAILKIQSEEGQRKAFSTCSSHLCVVGLFYGTAIVMYV 200

264 APKSRHPEEQQKVLFLILQFLS 285
    :|:   |.||.|  |:|:  ::.
201 GPRHGSPKEQKKYLLLFHSLFN 222
```

FIG. 6A

```
  1 MRPTWAGWLM.RWCCPPADTRCTTFL..CSGRAAPQQAGAPGHHPKYSLF 47
    | ||...::. ::::... ..:..::  :.  :  . .:..::  |.. .
  1 MGPTSVPLVKAHRSSVSDYVNYDIIVRHYNYTGKLNISADKENSIKLTSV 50

 48 PWIW.........KVRGLLPPPLTTDVVPVRLTLRANLSAADLLRGRGLPL 89
    .:|:         .: .||.   |... .... : :||. .||| | :..
 51 VFILICCFIILENIFVLLTIWKTKKFHRPMYYFIGNLALSDLLAGVAYTA 100

 90 PHVPHCPRTARLSLEGWFLRQGLLDTNLTASVATLLAIAVERHRSVMAVQ 139
    :  .: | :|. ..||||:| : ..|.||| .|||||:||. .:: :.
101 NLLLSGATTYKLTPAQWFLREGSMFVALSASVFSLLAIAIERYITMLKMK 150

140 LHSRLPRGRVVMLIVGVWVAALGLGLLPAHSWHCLCALDRSSRMAPLLSR 189
    ||.   . |:.:|| :.|| .| || ||  :|:|::||..:| : ||. :
151 LHNGSNNFRLFLLISACWVISLILGGLPIMGWNCISALSSCSTVLPLYHK 200
```

MATCH WITH FIG. 6B

FIG. 6B

MATCH WITH FIG. 5A

```
190 SYLAVWALSSLLVFLLMVAVYTRIFFYVRRRVQRMA..EHVSCHPRYRET 237
    |:  .  .     |::| :| :|.||:  .|| | .|:.  .::|  .| .|
201 HYILFCTTVFTLLLSIVILYCRIYSLVRTRSRRLTFRKNISKASRSSE. 249

238 TLSLVKTVVIILGAFVVCWTPGQVVLLLD.GLGCESCNVLALEKYFLLLA 286
    .:.|:|||:|:|:.|:.||.|  ::|||| |  ...|::|  ..|||:||
250 NVALLKTVIIVLSVFIACWAPLFILLLLDVGCKVKTCDILFRAEYFLVLA 299

287 EPTSLVNAAVYSCRDAEMRRTFRLLLRVPPPVHPRVCPLYILCPGRCQ 336
      .| .|: :|.  : ||||.| |                |:...:|.
300 VLNSGTNPIIYTLTNKEMRRAFIR.....IMSCCKCP 331

337 HSHHASRERPPTDGLHPLATLNYSGTR.....QATNPQPLMTCGC 376
    :.  |:: :.|.     :|.:::| .:      ::.||:.:|.:|
332 SGDSAGKFKRPI.....IAGMEFSRSKSDNSSHPQKDEGDNPETIMSSG. 375

377 SWLNPTS 383
      :|..|
376 .NVNSSS 381
```

```
  1 MNTTVMQGFNRSKRCPKDTRIVQLVFPALYTVVFLTGILLNTLALWVFVH 50
    :|.|  |. : |  |..:  |.| ::|.||.:||:.|||||.:. |:|..
  2 INSTSTQPPDES..CSQNLLITQQIIPVLYCMVFIAGILLNGVSGWIFFY 49

 51 IPSSSTFIIYLKNTLVADLIMTLMLPFKILSDSHLAPWQLRAFVCRFSSV 100
    :|||..|||||||.::||::|.| :|||||:|| |:||||..||||.|.|
 50 VPSSKSFIIYLKNIVIADFVMSLTFPFKILGDSGLGPWQLNVFVCRVSAV 99
101 IFYETMYVGIVLLGLIAFDRFLKIIRPLRNIFLKKPVWGKTVSIFIWFFW 150
    :|| .|||:||::|||.|||:.||::||:. |:... ::| :|:::|::
100 LFYVNMYVSIVFFGLISFDRYYKIVKPLWTSFIQSVSYSKLLSVIVWMLM 149

151 FFISLPNMILSNKEATPSSVKKCASLKGPLGLKWHQMVNNICQFIFWTVF 200
    ::::.:||:||.|... . .  || .||:.|| |||.  | |   |||.||
150 LLLAVPNIILTNQSVREVTQIKCIELKSELGRKWHKASNYIFVAIFWIVF 199
201 ILMLVFYVVIAKK.YMILIESPKVRTEKTTKSWKAKYLLSWLSSLCVL.. 247
    :|::|||..|.|| :.  :.|.: .|. ..||.:. : : ::  :|.:
200 LLLIVFYTAITKKIFKSHLKSSRNSTSVKKKSSRNIFSIVFVFFVCFVPY 249

248 .............LHFISPEFHILTVKPTIRLTVDC......KINCLLL. 277
                  |: :.. .||  ...: | :..      .| :::|
250 HIARIPYTKSQTEAHYSCQSKEILRYMKEFTLLLSAANVCLDPIIYFFLC 299

278 .........KKQLSFWQQ..LTFVWIP 293
            | ::.:..|  |.: :|.
300 QPFREILCKKLHIPLKAQNDLDISRIK 326
```

FIG. 7

```
  6 SSHCFYNDSFKYTLYGCMFSMVFVLGVISNCVAIYIFICVLKVRNETTTYMINLAMSDLL 65
    SS+C   DSFKYTLYGC+FSMVFVLG+I+NCVAIYIF   LKVRNETTTYM+NLA+SDLL
  3 SSNCSTEDSFKYTLYGCVFSMVFVLGLIANCVAIYIFTFTLKVRNETTTYMLNLAISDLL 62

 66 FVFTLPFRIFYFTTRNWPFGDLLCKISVMLFYTNMYGSILFLTCISVDRFLAIVYPFKSK 125
    FVFTLPFRI+YF  RNWPFGD+LCKISV LFYTNMYGSILFLTCISVDRFLAIV+PF+SK
 63 FVFTLPFRIYYFVVRNWPFGDVLCKISVTLFYTNMYGSILFLTCISVDRFLAIVHPFRSK 122

126 TLRTKRNAKIVCTGVWLTVIGGSAPAVFVQSTHSQGNNASEACFENFPEATWKTYLSRIV 185
    TLRTKRNA+IVC  VW+TV+ GS PA F QST+ Q N     CFENFPE+TWKTYLSRIV
123 TLRTKRNARIVCVAVWITVLAGSTPASFFQSTNRQNNTEQRTCFENFPESTWKTYLSRIV 182

186 IFIEIVGFFIPLILNVTCSSMVLKTLTKPVTLSRSKINKTKVLKMIFVHLIIFCFCFVPY 245
    IFIEIVGFFIPLILNVTCS+MVL+TL KP+TLSR+K++K KVLKMIFVHL+IFCFCFVPY
183 IFIEIVGFFIPLILNVTCSTMVLRTLNKPLTLSRNKLSKKKVLKMIFVHLVIFCFCFVPY 242

246 NINLILYSLVRTQTFVNCSVVAAVRTMYPITLCIAVSNCCFDPIVYYFTSDTNSEFNKNE 305
    NI LILYSL+RTQT++NCSVV AVRTMYP+TLCIAVSNCCFDPIVYYFTSDTNSE +K +
243 NITLILYSLMRTQTWINCSVVTAVRTMYPVTLCIAVSNCCFDPIVYYFTSDTNSELDKKQ 302

306 KL 307
    ++
303 QV 304
```

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/04079

A. CLASSIFICATION OF SUBJECT MATTER
IPC(6) :C07K 14/705, 16/28; C12N 15/12
US CL :435/6, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5, 24.31
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. 435/6, 69.1, 252.3, 320.1; 530/350, 388.22; 536/23.5, 24.31

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
NONE

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
NONE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim N .
A	NATURE, Vol. 361, issued 28 January 1993, K. Raming et al., "Cloning and expression of odorant receptors", pages 353-356, see Figure 1 on page 354.	1-32
A	Proceedings of the National Academy of the Sciences USA, Vol. 89, issued October 1992, P. Nef et al., "Spatial pattern of receptor expression in the olfactory epithelium", pages 8948-8952, see Figure 1 on page 8949.	1-32

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 09 JUNE 1995	Date of mailing of the international search report 19 JUN 1995
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer JOHN D. ULM Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

BNSDOCID: <WO___9630406A1_I_>

BAD ORIGINAL

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/04079

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☒ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/04079

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1.

Group I, claims 1-22 and 28-32, drawn to a nucleic acid encoding a putative receptor protein, the protein encoded thereby, and processes of use.

Group II, claim 23, drawn to an antibody which binds to a protein of Group I.

Group III, claims 24-27, drawn to a compound of undefined structure and chemical composition and methods of using that compound.

The inventions listed as Groups I, II and III do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The protein and nucleic acid of Group I, the antibody of Group II and the compound of Group III are chemically and structurally unrelated and do not share a common technical feature. Accordingly, the claims are not so linked by a special technical feature within the meaning of PCT Rule 13.2 so as to form a single inventive concept.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/04079

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	The Journal of Biological Chemistry, Vol. 265, No. 16, issued 05 June 1990, T. Hla et al., "An Abundant Transcript Induced in Differentiating Human Endothelial Cells Encodes a Polypeptide with Structural Similarities to G-protein-coupled Receptors", pages 9308-9313, see Figure 2 on page 9311.	1-32
A	Proceedings of the National Academy of the Sciences USA, Vol. 90, issued May 1993, L. Rohrer et al., "Cloning and characterization of a fourth human somatostatin receptor", pages 4196-4200, see Figure 1 on page 4197.	1-32
A	FEBS LETTERS, Vol. 298, issued February 1992, N. Iwai et al., "Identification of two subtypes in the rat type I angiotensin II receptor", pages 257-260, see Figure 1 on page 258.	1-32
A	SCIENCE, Vol. 244, issued 05 May 1989, F. Libert et al., "Selective Amplification and Cloning of Four New Members of the G Protein-Coupled Receptor Family", pages 569-572, see entire document.	1-32